A pilot evaluation study on benefits of a record linkage between a hospital diabetes database and the information systems within the NHS.

Weng, Chao

A pilot evaluation study on the benefits of a record linkage between a hospital diabetes database and information systems within the NHS

Chao Weng

The department of Medicine

Unit of Diabetes and Endocrinology

King's College, London

Guy's, King's and St Thomas' Schools of Medicine

St Thomas' Campus

A thesis submitted in part fulfilment of

the requirements for the degree of

DOCTOR OF PHILOSOPHY

of

UNIVERSITY OF LONDON

September 1999

Abstract

This study intends to establish a method to update the completeness and accuracy of patient's data on a hospital-based computerised diabetic patients records system, through record linkage techniques. The aim of this is to reach the targets of the St Vincent Declaration, by using IT technology, to continually monitor the outcome of diabetes care provided by the NHS.

Firstly, a record linkage was established between "Diabeta", (the computerised diabetes clinical records system in the Diabetes Unit of St Thomas' Hospital) and the National Health Services Central Register (NHSCR) (in the Office for National Statistics (ONS)). This is expected to provide an update of the "Diabeta" database in respect to patients' vital status (alive/dead) and used to assess premature mortality in diabetes, as it is acknowledged that many deaths currently occurring within the population, under the care of the Diabetes Unit, are not notified. Secondly, the possibility of linking "Diabeta" with other NHS information systems has been evaluated based on a sample of records extracted from "Diabeta". This was done in order to obtain information about hospitalisation, emergency care, laboratory reports and home-care etc. This was then used to evaluate the resources and assessment of the cost of diabetes care, considering that data on "Diabeta" is primarily only about the long-term out-patient care of diabetic patients.

This study shows that a record linkage between "Diabeta" and the NHSCR gave an update of 91% of the records in "Diabeta", and that most of the patients' vital status would be updated as a result of the linkage, the benefits of which are considerable. It was demonstrated also that at that time 86% of deceased diabetic patient's data had not been notified to "Diabeta". "Diabetes Mellitus" was recorded as a cause of death in only 36% of death certificates of people known to have diabetes. Based on the updated vital status in "Diabeta", meaningful mortality studies in the diabetic population were conducted. The author recommends that by linking "Diabeta" with other relevant information systems, a complete picture of care provided to patients with diabetes will be obtained on "Diabeta", thus enabling the construction of an outcome study (clinical/economical) in diabetes care.

Acknowledgements

I would like to thank Professor P H Sönksen for providing valuable advice, guidelines and encouragement during the project, and also Dr D V Coppini for his supervision and support during my research.

During the course of my PhD, the "Diabeta" team in the Department of Medicine at St Thomas' Hospital has continuously supported my study. Mr Steve Carey and Mr Andy Broome have helped to extract data from "Diabeta" for the 'flagging' study in the NHSCR.

My appreciation also goes to the medical research staff in the NHSCR who spent their time tracing our records on the NHSCR database and providing us with updated patient's information.

In the Department of Public Health at UMDS (now known as GKT), Dr Mark Shaw and Mr Peter Aspinall have provided statistical advice, census data and social deprivation data. Dr Philip Young, senior statistician also contributed many ideas on survival analysis. Mr Mike Farrel, system assistant in a MapInfo company, provided technical support and helped me to apply Geographic Information System (GIS) techniques. I am also grateful to various system managers/analysts of the IT departments who kindly contributed their time to this study. These IT departments are at St Thomas' Hospital; the Lambeth, Southwark and Lewisham Health Authorities; various general practices in LSL area; Lewisham Health Care; Lewisham Council; Prescription Pricing Authority; and Nation Wide Clearing Services.

I am very grateful to UMDS and Roche in UK who sponsored my thesis and enabled me to complete the study without financial worries.

I also acknowledge all the help, support and encouragement of my friends and family throughout my studies.

Table of Contents

List of Abbreviations

A+E	Accident and Emergency
ACF	Accelerated Failure Time Model
BDA	British Diabetes Association
BMF	British Medical Formulation
CAN	Cancer
CHRIS	Central Health Register Information System
CIS	Clinical Information System
CMDS	Contract Minimum Data Set
CMMS	Case Mix Management System
CVD	Cerebrovascular disease
DEDC	Diabetes and Endocrine Day Centre
DKA	Diabetes Ketoacidosis Coma
DNA	Did Not Attend
DoH	Department of Health
E & W	England and Wales
EDI	Electronic Data Interchange
EPR	Electronic Patients Records
ESRF	End Stage Renal Failure
FHS CU	Family Health Services Computer Unit
FHSA	Family Health Services Authority (now known as Health Authority (HA))
GIS	Geographic Information System
GKT	Guy's, King's and St Thomas' Medical and Dental School
GUY	Guy's Hospital
HTML	Hypertext Markup Language
ICD-9	International Classification of Disease, reversion 9
IDDM	Insulin Dependent Diabetes Mellitus
IHD	Ischaemic Heart Disease
LAN	Local Area Network
LCS	Local Clearing Services
LSL	Lambeth Southwark and Lewisham

NHSCR	National Health Services Central Register
NHSnet	NHS national network
NIDDM	Non Insulin Dependent Diabetes Mellitus
NWCS	NHS Wide Clearing Services
ONS	The Office for National Statistics (previously known as The Office of Population Census & Surveys
OP	Out-patient
OPCS	Office of Population Census & Surveys (now known as Office for National Statistics)
PAS	Patient Administration System
PCG	Primary Care Group
PPA	Prescription Pricing Authority
PVD	Peripheral Vascular Disease
RCGP	Royal College of General Practitioners
RMI	Resource Management Initiative
RRS	Report Result System
SMR	Standardised Mortality Ratio
SMRLL	Confidence interval for SMR Lower Limit
SMRUL	Confidence interval for SMR Upper Limit
SMTP	Simple Message Transfer Profile
STH	St Thomas' Hospital
UN/EDIFACT	United Nations/Electronic Data Interchange for Administration Commerce & Transport
UPA	Under Privileged Area
WAN	Wide Area Network
XML	eXtensible Markup Language

List of Figures

Chapter 10

Chapter 11

Chapter 12

Chapter 13

Chapter 14

List of tables

List of Appendixes

Preface

Diabetes Mellitus

Diabetes Mellitus is a syndrome characterised by hyperglycaemia due to absolute or relative deficiency of insulin. It is classified as primary diabetes and diabetes secondary to other pathology. Primary diabetes is classified as Type 1 and Type 2 diabetes mellitus.

Complications of Diabetes

Diabetes is a major cause of premature mortality, morbidity and reduced quality of life. It is a chronic, systematic disease that may result in long-term morbid manifestations related to both micro-vascular (e.g. nephropathy and retinopathy) and macro-vascular complications (e.g. cerebrovascular disease, and peripheral vascular disease). The probability of developing these complications is related to the duration of disease and the degree of metabolic controls (Currie CJ et al 1996).

These complications have very serious consequences and may lead to amputation, renal failure or blindness.

The complications of diabetes contribute heavily to the burden of health care and consume large of amounts of resources. Foot problems, renal failure and cardiovascular disease account for a high proportion of hospital admissions, leading to considerable disability, and considerable premature mortality.

The prevalence of diabetes in the UK is about 1-2% (Nabarro JDN 1988) and 4-5% of total health care expenditure is devoted to the care of people with diabetes (Currie CJ et al 1997). Costs are dominated by in-patient care for diabetes complications.

With adequate screening (eye screening, urine test, neuropathy screening) and intervention provided by the large range of clinical professionals available, most of the complications can be, and the costs of services for people with diabetes can be reduced.

In order to raise awareness on the importance of undertaking comprehensive and systematic action to prevent diabetes complications, the World Health Organisation (European Branch), in partnership with the International Diabetes Federation (European Branch), organised a conference in Italy (1989) which resolved to tackle fundamental issues, summarised in the "St. Vincent Declaration". The five year target of the St Vincent Declaration is to reduce blindness and end-stage diabetic renal failure by at least one third, the rate of limb amputations for diabetic gangrene by half, and to significantly reduce morbidity and mortality from coronary heart disease in diabetic patients.

The St Vincent Declaration and the use of IT in diabetes care

One aspect tackled in the St. Vincent Declaration was the role of IT in supporting the management of patients with diabetes acknowledging that in the fast growing field of IT, greatly improved data handling and management of patients with diabetes have been accomplished over the last two decades.

In a European Conference held in Budapest (March 1992 - a follow-up recently to the implementation of the St Vincent Declaration), participants addressed the importance of quality assurance in diabetes care and discussed the use of IT to improve it. During the conference, delegates repeatedly stressed the importance of the availability of information in helping to achieve the goals of the Declaration. It was acknowledged that the quality of diabetes care remains below what is desirable, and the standard of care actually achieved (process or outcome), is rarely measured (RJ Young et al 1999). Although many data items are routinely collected as part of a clinical review, they are not always available in a standardised form (e.g. standardised baseline and outcome measures), and not easily collated for analysis. Moreover, all the process of treatment and its outcomes, for each individual patient, is not always recorded in a single system. This is due to the high mobility of those patients living in an inner city and to the variation of diabetes services available. There is increasing demand for better communication between computerised systems handling information on diabetic patients. This helps to draw together computerised records belonging to the same individual, enabling comprehensive monitoring of the outcome of services. The quality

of outcome research relies on the accuracy and completeness of updated data, accumulated in a central information system.

Outcomes on health care can be measured in clinical, behavioural, economic and political aspects. The clinical outcome can be measured by biochemical parameters (e.g. HbA1), physiological parameters (e.g. visual acuity, nerve conduction velocity) and pathologic parameters (e.g. blindness, end stage renal failure, myocardial infarction, amputations and mortality). Behavioural outcome is related to the quality of life (e.g. quality of well-being). The effectiveness of health care in economic terms can be measured by direct cost (hospitalisation, out-patient care, diagnostic testing, therapeutic intervention) and indirect cost (e.g. lost productivity and rehabilitation etc.). By cross-referencing economic outcome with the clinical/behaviour outcome of health care, evidence for political decision making can be produced. These measures are set to ensure that all patients with diabetes obtain equal care, and to achieve the best outcome of health care within a given community by using minimum NHS resources. In order to do so, a routine information system should record measures of outcome (clinical, behavioural and economic) for different conditions together with simple prognostic information relating to the characteristics of patients and to health care intervention.

The Declaration stated that it was necessary to establish monitoring and control systems, using "state of the art" information technology, for quality assurance of diabetes. Data needs to be standardised as recorded in the computerised diabetes records system, and measured in a timely-fashion (British Diabetic Association 1995). Systems must be capable of continually showing the effectiveness of treatment of diabetes and its complications.

In order to achieve this, information systems related to care provided to diabetic patients, needs to be completed and then linked. The challenge is to apply the concept throughout Europe. Such links can facilitate the electronic interchange of patient data ensuring automatic import of updated data between the systems involved, producing a complete picture of care provided to patients with diabetes in a central system.

EURODIABETA is one of the largest (16 partners, 3 sub-contractors, from 7 EC member states) Advance Informatics in Medicine (AIM) funded projects bringing European experts together to perform a feasibility study on the production of a (computer assisted) chronic health care environment to support diabetes care. The intention is to integrate clinical care, diet therapy, medical record management and communication and co-ordination with other providers and to develop an effective chronic health system.

Following these general considerations, this study will examine the possibility and benefits of a record linkage between "Diabeta" and some other information systems within the NHS.

This thesis will show the findings from the present research, in four sections: **Introduction, Methodology, Results, and Discussion & Recommendations**. Section One contains three chapters, the first of which introduces the background, functions, strengths and weaknesses of "Diabeta".

Section Two contains 5 chapters describing the research activities performed, the patients used in the pilot study and methodology of linking "Diabeta" to corresponding NHS information systems. Various statistical analysis techniques and geographic information systems are also explained.

In Section Three, five chapters illustrate the results of records linkages based on two cohorts of patients and the benefits in terms of outcome of diabetes care studies.

Section Four (2 chapters), examines errors occurring in record linkages and validates the strength of an epidemiology study. Further system developments on "Diabeta" are recommended.

Section One

Introduction

In order to assess the suitability of developing record linkages for improving the quality of data recorded in a local computerised database by other invariable databases, the following questions have to be addressed:

1. What data items need to be updated in the local database?
2. To what extent is it procurable from the other valid databases?
3. Will the benefits that result from the update of the local database counteract the costs of implementing this record linkage?

In this section the suitability of a records linkage between "Diabeta" and other information systems will be discussed.

Chapter 1

Challenges to the computerised diabetes clinical records system

1.1 "Diabeta" - the computerised clinical records system at St Thomas' Hospital

"Diabeta" is a computerised medical record system, designed to improve the process of management and care of diabetic out-patients at St Thomas' Hospital. It consists of a suite of software written in UNIFACE 4 GL and runs as a client/server system with Window NT, 95 & 3.1 clients against Oracle v. 7 on a HP-UX server. It can be accessed by all medical staff in the Diabetes Unit, medical clerks and laboratory technicians.

At present, "Diabeta" holds 10 653 diabetic patients' records and 4113 records of endocrine patients, with about 400 new cases of diabetes entered into "Diabeta" each year.

1. 1. 1 Data items collected in the records in the "Diabeta" Database

Each patient record includes a maximum of 183 data items covering the following topics:

Registration

-identification details of patient
-demographic details
-GP code, name and address
-source of referral

First referral

-diabetic history
-other medical history
-family history of diabetes
-obstetric history
-social history (including smoking and alcohol)
-medication history
-general examination
-eye examination
-lower limb examination
-problem list management details
-medication list
-diet details

-investigations

-referrals

-date of the next appointment

Follow-up visit

-date of visit

-initials of person conducting consultation

-measures of control

-weight

-random blood sugar

-haemoglobin A_1

-complication/risk factors

-blood pressure

-visual acuity

-fundus examination

-updated problem list and treatment

These data items are entered into the system mainly by doctors during consultation of the patient's first visit (**first visit clinics)**. During follow-up visits (**follow-up clinics**) only a subset of maximum 30 data items is entered which are most likely to change in time (time-related data).

Follow-up clinics are scheduled according to the doctors' recommendation, taking into consideration the type of diabetes, patient's age and metabolic control. Within the follow-up module, visual acuity, urine analysis and blood pressure are checked only once a year. These 'complete' follow-up sets are called **annual reviews** so that they can be distinguished from the **routine follow-ups**.

Special clinics are held for pregnant women with diabetes where specific pregnancy data is collected, in addition to the established data set.

Diabetes nurses and specialists also enter some patient data when seeing patients. Other professionals such as dieticians, chiropodists or eye specialists may also enter specific data.

Clinic clerks enter patient registration details.

This routine entry of medical data has formed a cumulative, accessible, revisable, analysable and amendable database, allowing electronic manipulation.

The data entry process is organised by a structured questionnaire that includes branches to exclude questions irrelevant to a particular patient. The possible routes through the questions range from 77 to a maximum 176. This questionnaire is presented along a series of screens. These screens are termed 'pages', each page covering only one topic: "**Registration Page, Examination Page, Problems and Operation Page, and Management Page"**. Data is entered into the system by typing on a keyboard in response to the questions displayed on these screens.

The most relevant screen to this study is the "**Registration Page**" and "**Problems and Operation Page**" as it displays a sectional view of the structured questionnaire, specifically allocated to the collection of patients' demographic and diagnosis details. This is illustrated in Figure 1.1 and Figure 1.2

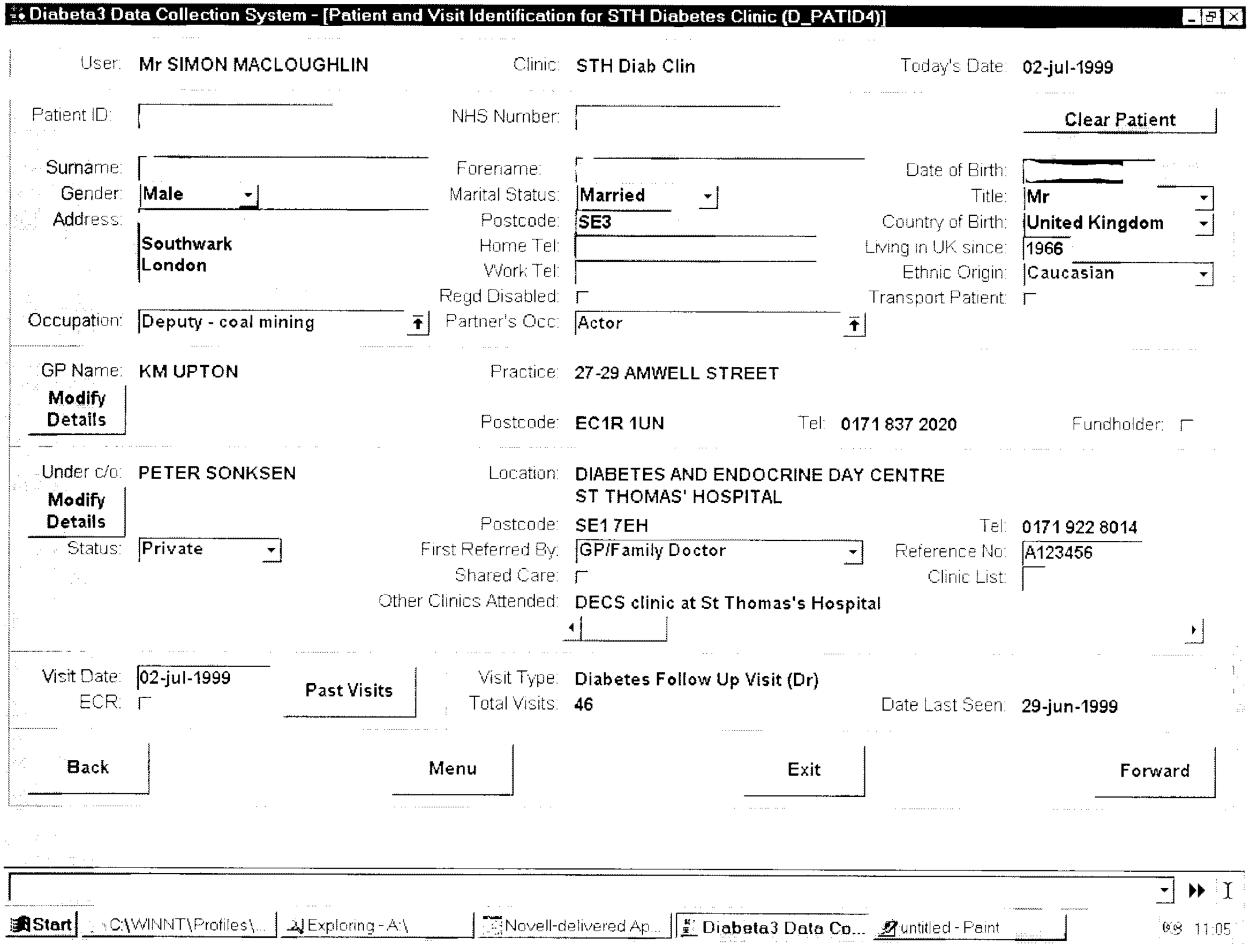


Figure 1.1 'Registration Page' in "Diabeta"

Diabeta3 Data Collection System - [Problems/Diagnoses List and Procedures/Operations List (D_DANDP1)]

User: Mr SIMON MACLOUGHLIN
Patient:

Age: 61 (16-feb-1938)

Today's Date: 16-aug-1999
Visit Date: 17-jun-1999

Problems/Diagnoses List and Procedures/Operations List (D_DANDP1)

Seq	Problem/Diagnosis	Comments	Date Diagnosed	Status
1	diabetes mellitus		Feb/1956	active
2	charcot joint r foot	Swollen & Hot - great in 6/99	Dec/1995	inactive
3	background retinopathy	Dots & Blots	1974	active
4	peripheral neuropathy	Ataxia & Dodgy Feet	Oct/1984	active
5	abdo pain & vomiting	Fungal Oesophagitis	Feb/1995	inactive
6	Haematemesis & Melaena	? Mallory Weiss	Mar/1994	inactive
7	Ulcer left foot	Healing Slowly 94/95/96 - healed in 6/99	Mar/1995	inactive
8	infected heel	At Kings	24-jul-1997	inactive
9	depression	Long Term Medication	1975	inactive
10	Diarrhoea	Admitted and needed i/v fluids & antibiot	08-oct-1998	inactive

Add problems | Delete the highlighted problem

Seq	Procedure/Operation	Comments	Date performed
1	*cataract & lens replacement	Left Eye; R Eye 1/94	Sep/1993
2	*carpal tunnel decompression	Right & Left	1981
3	*caesarian section		1960
4	Photocoagulation (Right Eye)		?/1989
5	Photocoagulation (Left Eye)		?/1989

Add procedures | Delete the highlighted procedure

Back | Menu | Exit | Forward

Figure 1.2 The 'Problems and Operation Page" in "Diabeta".

The **patient's name** (surname, forenames), **CB** number (hospital number) [1] and **PN**[2] are patient identifiers. Each of these identifiers is sufficient to retrieve a patient's medical record from the entire system database.

If known, the PN is the quickest way access a patient's record. "Diabeta" holds the patient's NHS number[3] as well, although this is not routinely entered. This is because this number is rarely known by the patient and there is no quick means of tracing this number when the patient is registered with the DEDC. The unavailability of the patient's NHS number in "Diabeta" is becoming a challenge when considering future electronic linkages of this local system as against more central information systems within the NHS Wide Network. Currently, the new NHS number (an unambiguous number with 10-digits) provides the main means of electronic data interchange among NHS computer systems at the national level.

1. 1. 2 Development of the "Diabeta" System

The Diabetes Unit of St Thomas' Hospital uses information technology to enhance the efficiency and usefulness of medical records, and has a 28-year history. Over this period, computer analysts and clinicians have made great efforts to achieve the current operational records system in "Diabeta".

The initial prototype of the current "Diabeta", developed in 1982, has been gradually and progressively enhanced to expand its role within the 'diabetic team'. In order to support some aspects of diabetic in-patients, it also takes in other clinics e.g. general medical, or lipid clinics and its development stages are summarised as follows:

The 70's

In 1971, a team of clinicians and computer analysts from St Thomas' and King's College Hospital started to attend weekly workshops in order to decide upon standardised, structured questionnaires for the data set to be collected for diabetic out-patients.

By 1973 this data set was defined and reached maturity and stability. Subsequently, paper forms displaying the structured questionnaires were routinely used to collect the clinical details of all newly diagnosed diabetic patients and then transferred onto punched cards with the batch entered onto St Thomas' Hospital main computer.

---

[1]CB is a Hospital Central Bureau Number

[2] PN is "Diabeta" Patient Number that is allocated to each patient by the system.

[3]NHS number is an integrated numerical identifier arranged to each person when registering with National Health Service.

With the introduction of the haemoglobin A_1 ($HbA1_1$) measurement as a routine method for assessing diabetes control over the last three months, and also with the acquisition of a Biothesiometer, the standard questionnaires that had been used for the data collection were updated accordingly. Since 1979, HbA_1 measurement and Biothesiometer readings have been also recorded in paper forms.

The 80's

At the beginning of the "Diabeta 1" project, the batch procedure was used for entering medical records because adequate hardware for interactive entry was not affordable for the Diabetic Unit.

The 'Microcomputer Revolution' re-activated the research activities in the Diabetes Unit of St Thomas' Hospital and toward the development of a prototype with interactive data entry. A single-user microcomputer was installed in the Endocrinology Ward together with a printer (affordable at that time) by 1982. The microcomputer was (optionally) used by doctors in the diabetic clinic to interactively enter information about newly diagnosed patients attending a first visit clinic at St Thomas' Hospital.

Subsequently, with the provision of multi-user microcomputers, the interactive data entry was expanded for all diabetic patients and it became possible for every doctor in the diabetic clinic to enter this data during medical consultations. The final prototype of the "Diabeta 1" (written in APL and running on a NCR, Tower, UNIX-based system) was completed in 1986.

The 90' s

The prototype "Diabeta 1" was continuously updated. "Diabeta 2" was developed in 1986, using Database "Sculptor" in order to link rule-based "expert system". It was an extension of "Diabeta 1", covering additional data collection for "expert system" and only used in Brighton and Eastbourne. The 3rd generation Clinical Information System (CIS) -"Diabeta 3"- was developed and went "live" at St Thomas' Hospital on 1st of January 1998. It updated and merged functions of "Diabeta 1" and "Diabeta 2" into a common system, adding "Open System" and "Configuration". "Diabeta 3" is extremely versatile and provides focused, customised clinical data collection. It also has a sophisticated clinical content configuration (e.g. visit type, examination types, screen and report library and clinical terms). "Diabeta 3" is still developing in respect to retinal image handling and web-based version.

1. 1. 3 The Endocrinology and Diabetes Day Centre

The progress in integrating the activities of all the members of the 'diabetic care team' in a coherent and efficient way for better patient management was achieved due to the development of a computerised medical records system as described. It was also due to the organisational reforms finalised within the setting of the current Diabetes and Endocrinology Day Centre (DEDC).

The centre was partially operational from October 1989, offering limited services only, and became fully operational in February 1990.

The DEDC is located on the 6th floor at St Thomas' Hospital and includes the following facilities:

Waiting area, reception
3 consulting and examination rooms
Laboratory test area
Dietician's room and kitchen
Group education area
2 diabetes specialist nurses' rooms
4 out-patients' day beds
Secretarial office
Chiropody room
Diabetic eye compilation screening room
Seminar room
Computer manager's office
Sanitary points

The DEDC is a self-contained unit, which requires a minimum of support services from the rest of the hospital.

1.1.4 The out-patient attending the DEDC

Most of the patients coming to the DEDC are referred by GPs who write referral letters to the consultants requesting appointments. Sometimes patients may be referred by Accident and Emergency, child psychiatrists, community physicians, surgeons, or from other hospitals. Rarely patients may refer themselves. In each case, a letter or report from another unit will be handed in by patient relative or by the patient himself or herself at the reception of the DEDC.

All patients attending their **first visit clinic** in the DEDC are referred to as **'new patients'** and if they have not been seen in the hospital before, a hospital number is allocated by the clerks working at reception.

On their first visit, patients have a routine blood and urine test performed by a clinic nurse, taking approximately 15 minutes. Some patients may then be referred for specialist treatment either by the specialist nurses, chiropodist, dietician or the diabetic eye complication specialist. The doctor's clinic visit is about an hour long and clinical information (e.g. full medical history) is entered into "Diabeta" during the consultation.

For a newly diagnosed insulin dependent diabetic, after his/her first visit with the doctor, the patient is asked to participate in a series of encounters referred to as **work-up visits**. Initially the clinic nurse will measure the blood glucose concentration and the patient then sees the diabetic specialist nurse (possibly on the same day but usually on another agreed day). Appointments will be daily for the first week with each visit lasting 30 minutes to 1 hour and then once or twice in the second week for 30 minutes or maybe once a week for 30 minutes for the next 2-3 months (It varies according to the individual patient). There is also a telephone contact available with the specialist nurses during the day. The patient is also seen by the dietician for 45 minutes and again at 2 and 4 weeks for thirty minutes. In those NIDDM patients who do not need to be taught how to administer insulin, explanation and discussion of their condition performed on a group of ten bases. There are two sessions for group teaching at intervals of one month, with the first lasting 2-2.5 hours. The specialist nurse teaching session is 1-1.5 hours long and that of the dietician about 45 minutes long. If patients have special difficulties or language problems, then they will be serviced by the specialist nurse individually on at least two occasions.

Once the initial work-up visits are completed and the patient's treatment stabilised, he/she then attends the regular follow-up visits that occur at six monthly intervals (although emergency treatment is given if necessary). When the follow-up patient attends the DEDC, routine tests are performed by clinic nurses within 5 minutes of the doctor consultation. The follow-up consultation with the doctor takes about 15 minutes. At follow-up semi-automatically generated letters printed out by doctors and handed to patients for their GPs.

If patients fail to attend their booked clinical appointment, the clerk will enter 'DNA' ('did not attend') on the item of 'doctor seen last' (within the follow-up module) in their terminal installed with "Diabeta" software, and also notify the doctor. Patients are usually sent another appointment after the 1st 'DNA'. In case of a subsequent 'DNA' a letter is written to the patient's GP.

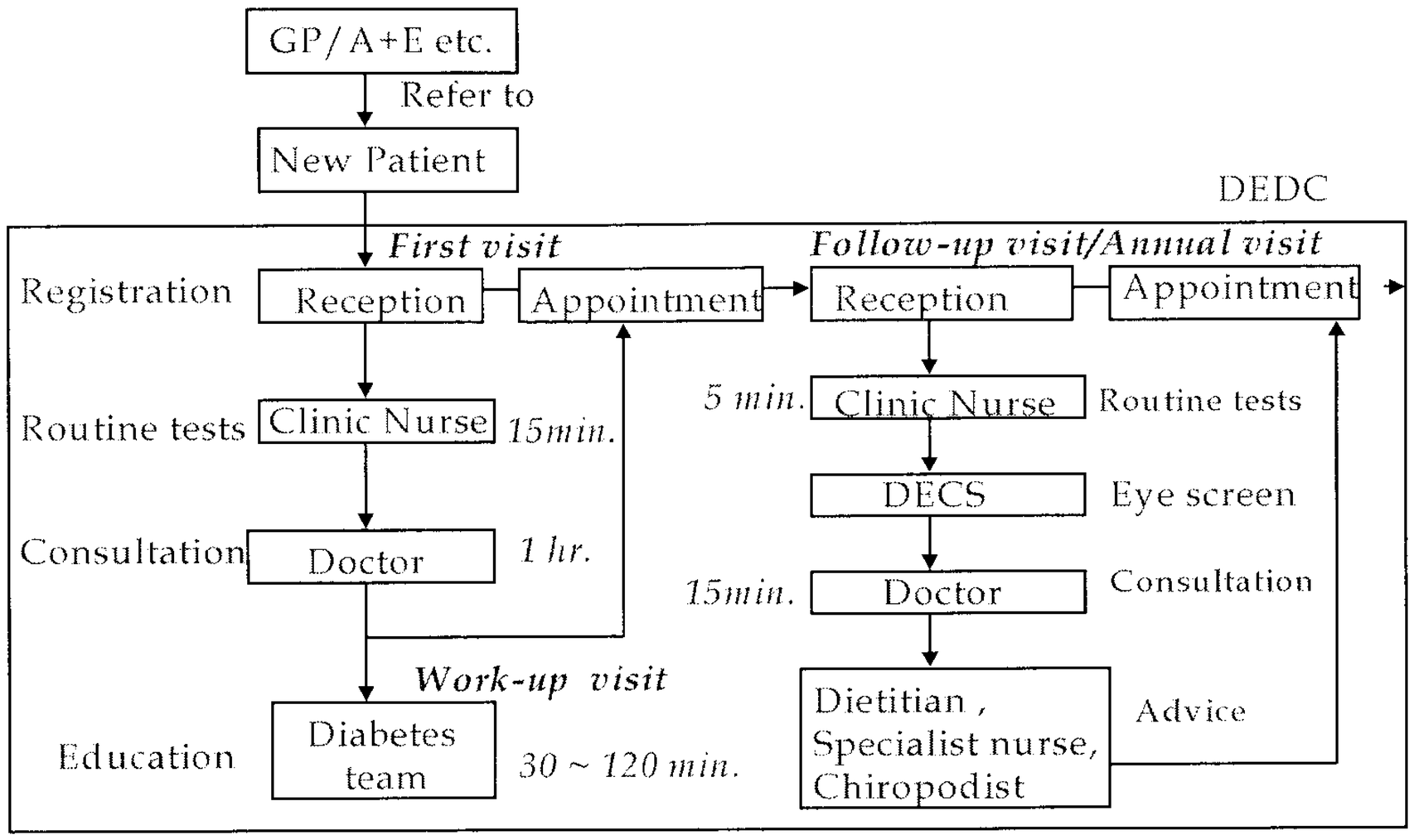


Figure 1.3 Diagram of out-patient clinical activity flow in the Diabetes and Endocrine Day Centre.

"Diabeta", which routinely collects a core clinical data set, has proved to be effective in the management - registration, recall and regular reviews of diabetes, whilst also providing in an invaluable tool for research. However, the quality of data in "Diabeta" is still a challenge for epidemiological studies.

1. 1. 5 Challenges to the accuracy of patient demographic data held on the "Diabeta" system

The patient's demographics may change between two subsequent visits. These changes cannot be updated, unless the patients, their relatives or their GP notify the diabetes clinic at St Thomas' Hospital whenever it occurred. The accuracy of the information on out-patients' vital status (alive/deceased) has always been acknowledged to be poor, and notification of deaths to the clinic is notoriously incomplete. Nevertheless the changes on patient's current address or/and GP details cannot be updated until patients have been seen in the diabetes clinic. It is also known that maintaining accurate demographic data on patients in "Diabeta" is especially difficult within an inner city where up to 30% of population's registration data changes each year.

The erroneous data may include

- patient's current address or/and
- details of patient's GP: address, code
- patient's vital status (dead/alive)

One of the most important roles of "Diabeta" has been to support research activities in the Diabetes Unit. Datascan, the section of "Diabeta" specifically designed for data retrieval and analysis, has proved to be very powerful but the accuracy of the results may be impaired if the database cannot be updated as soon as changes in patient's information occurs. Accurate analysis of mortality and the ability to perform longitudinal studies in the diabetes clinic population is impossible unless the patients' vital status is continually updated.

The inaccuracy of a patient's address and/or GP details is important only when certain patients are to be contacted for a specific research study, in this case the researcher has no means to trace that patient, and relies only on feedback from the patient .

1.1.6 Challenges to the completeness of patient clinical activities data held on the "Diabeta" system

Diabetes is a long-term, chronic disease. Complications can lead to a physical handicap, poor quality of life and premature death. With appropriate intervention, these complications can be prevented and treated. Decision-making about intervention needs a collaborative approach to detect complications at an early age and to treat them when they first occur. The main components of current diabetes services are hospital-based diabetes teams (like DEDC at St Thomas' Hospital), the primary care team (general practitioner with appropriate nursing and administrative support) and other community support (e.g. eye screening, chiropody etc.). Other health services, such as hospital in-patient clinics and Accident and Emergency services are also heavily involved in the treatment of complications of diabetes (e.g. diabetic ketoacidosis).

Once the target for prevention of complications is clear, the clinician is in a better position to specify interventions, services or treatments. In some cases this may be a single intervention, for example, a limb amputation, but in many cases a whole range of interventions (e.g. weight control, metabolic control, drug treatment, operation etc.) is involved. In order to know which intervention is more cost-effective for improving the quality of patient's life, a valid outcome study needs to be conducted with a system which contains the following four categories of patient data:

- Patient demographic and clinical data before intervention (e.g. baseline data)

- Intervention activity (eye screening etc.)
- A whole range of clinical activity

(admission, attendance, laboratory tests, prescriptions, etc.)

- Outcome data (metabolic controls, mortality, morbidity etc.) after intervention

Since the DEDC is a self-contained diabetes clinic, "Diabeta" is mainly designed for collecting out-patient clinical information within the DEDC. Information related to treatments in other settings (such as hospital in-patient clinics, Accident and Emergency, primary care etc.) cannot be collected into "Diabeta" automatically. This information can however be entered into "Diabeta" into the field "problem list and treatment" on **Problem and Operation Page** (or patients' paper file) if patients/ GPs/ other health care professionals report such information. Again, the completeness of these clinical activity data collected into "Diabeta" is known to be insufficient.

Since the quantity of care provide from outside the DEDC to these patients remains uncertain, the overall cost of utilisation of health services related to diabetes or its complications cannot rely on the "Diabeta" database alone. It is also impossible to identify which intervention has affected the patient's outcome.

1. 1. 7 Possible solutions

Although "Diabeta" has improved the accuracy and completeness of medical data, it is still insufficient. Establishing a better communications system between "Diabeta" and another reliable information sources would be desirable.

The "Diabeta" database can be updated in respect to the patients' vital status if the system is linked to the NHS Central Registrar (NHSCR), which is a division of the Office of Population Census and Surveys (OPCS) (now known as the Office for National Statistics (ONS)). The NHSCR holds a standard set of demographic data for all the population registered with a NHS doctor in England and Wales and contains a record of all death notifications.

The NHSCR data set does not include the patient's address or patient's GP details, but does include updated vital status, and the code of the Family Health Services Authority (FHSA)(now known as the Health Authority) which the patient is currently registered with.

The "Diabeta" database can potentially show the clinical information on emergency care, hospitalisation, laboratory reports and home care if linked with such sources.

The "Diabeta" database can be updated in the following ways:

–If it is linked with the data set in the NHSCR, this would continually update patients' vital status.

–If it is linked with the FHSA (at the St Thomas' Hospital catchment area), patient's demographic data would be continually updated.

- If it is linked with data set in the Accident and Emergency at St Thomas' Hospital, it can continually record A+E attendance of diabetic patients.

- If it is linked with the contracting data set in the main hospitals (or Health Authority, Department of Health), all diabetes related in-patient activities (e.g. vascular surgery, coronary care) could be continually recorded and updated.

- If it is linked with the GP data set, this could continually update attendance in general practice or community care and other health care settings (including drug dispensing).

- If it is linked with data set in the laboratory database at St Thomas' Hospital, relevant related clinical tests could be continually collected into the database.

The linkage between "Diabeta" and these information systems will be discussed in further detail in subsequent chapters.

Chapter 2

Considerations of linking existing information systems within the NHS to "Diabeta"

2.1. NHS registration systems considered for record linkages with "Diabeta"

In England and Wales the registration of population within the NHS relies on *the National Health Services Central Registrar* (NHSCR) and also on ninety-eight *Family Health Services Authorities* (FHSAs) [now known as Health Authorities (HA) and 60 of them in England and Wales], plus *Local Registrars* and *General Practices*.

Local Registrars are the corner stone of population registration in the NHS (The population born in England and Wales). Each Local Registrar is responsible for the registration of births and deaths of local residents, and for delivering notifications to the NHSCR.

Each **General Practice** provides health care services to all persons living in the geographical area of that practice (on a request basis) and report to the local FHSA on all new patients encountered.

FHSAs have a key role in building accurate, up to date and accessible information about potential users in the NHS. Each FHSA is responsible for registering individuals within its boundary, and to notify new ones to the NHSCR (the names of the newly registered patients associated with the code of that FHSA).

The **NHSCR** collects information received from all FHSAs, Local Registrars and other public service units (i.e. ports of embarkation, Department of Health, Home Office etc.) or members of the public, and integrates it in a central database. The NHSCR also disseminates the information received from Local Registrars (notification of deaths) and from other public service units (i.e. information on embarkation, long-stay psychiatric, prisoners and H.M. Armed Forces etc.) to the appropriate FHSAs.

2.1.1 The population registration process within NHS

I. Registration in the NHS of all new-born children in England and Wales

Every new-born child in England and Wales is recorded in the administration software of that labour ward. When the baby is delivered, the birth certificate is completed by a midwife. The new-born's parents would be advised by their local nursing staff to visit the Local Registrar for registration of birth within 6 weeks of the date of birth. A

receipt on notification of birth is also sent by the hospital to the Local Registrar. The NHS number of the new-born child is issued by the local Registrar's officer when the parents register the baby and will also forward notification of birth together with the NHS number to NHSCR. In conclusion, all new-borns in England and Wales are registered with the NHSCR.

II. Registration in the NHS of persons not born in England and Wales

General Practitioners send a pre-defined set of data on the population addressed to them to a local FHSA. For patients notified by local General Practices, FHSAs select those who are not already registered within the NHS (people not born in England and Wales) and will advise the NHSCR upon allocation of NHS numbers for these patients.

III. Deletion from local FHSAs and GP lists of some categories of patients registered with the NHS

The NHSCR informs each local FHSA when its patients 'exit' from the NHS (such as deaths, embarkation, long-stay institutions, psychiatric, prisoners or those that have joined H.M. Armed Forces). Subsequently, local FHSAs will delete these patients from their lists and will also advise the appropriate General Practices to do so.

The population registration process within the NHS is illustrated in Figure 2.1.

The existing NHS Number has various formats (there are approximately 26 formats) and are related to personal details (e.g. date of birth). This is not suitable for the exchanging (or sharing of) information about patients within current NHS information systems. In order to identify each person in an unambiguous and unique way, the NHS executive has decided to replace the existing NHS number. The new NHS number with 10-digits was put into operation for all new-borns after 4 December 1995 by Local Registrars.

In order to replace the existing NHS number, the NHSCR has developed CR software to generate new NHS numbers to handle both old and new NHS numbers. This will produce a tape for each FHSA to load new NHS numbers into FHSA registers. GP systems and other systems will load the new NHS Numbers through the link with the corresponding FHSA. Secondary care settings (hospitals) can update the New NHS number through a 'tracing service' in the NHSCR (NHS Number & Tracing Services (AR) Programme, 1997).

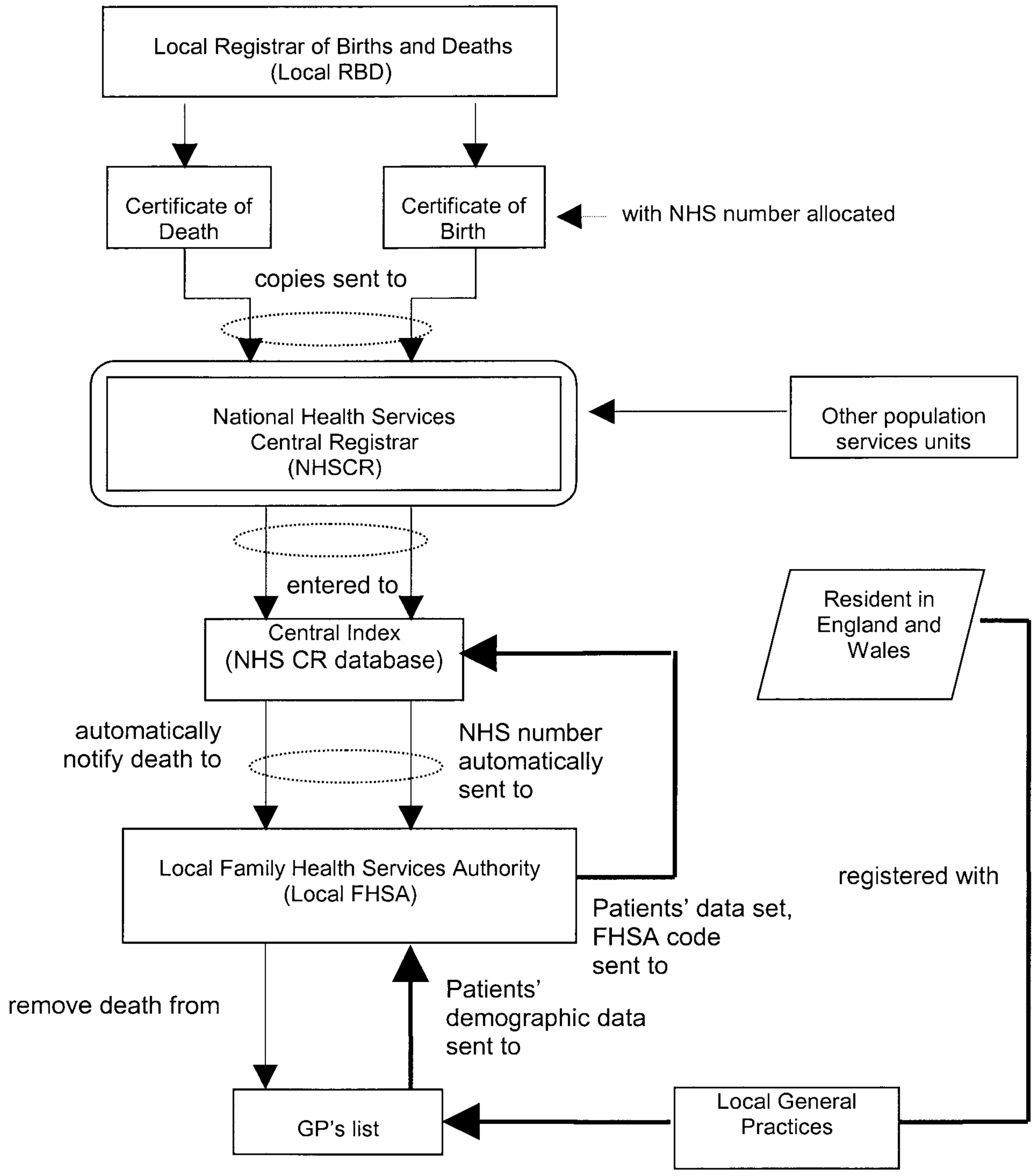


Figure 2.1 Population registration process within the NHS

2.1. 2 Patient demographic data held in main NHS registration systems

I. Patient Registration Databases in local FHSAs

Every FHSA has a computerised patient registration system (UMPS DBMS) recorded and run under the UNIX platform. The FHSA patient registration system (known as Exeter system) contains demographic details on all people living within the boundaries of that FHSA, detailed as follows:

- NHS number
- GP code
- Patient's name (surname, forename, previous name)
- Patient's address
- If patients register with a different FHSA, the previous FHSA will hold the date, reason for transfer and the code of the new FHSA.
- The code of previous FHSA, date of transfer from that FHSA.

This information is being constantly updated by other central databases (e.g. NHSCR) and as such, forms a large and accurate area database. The main function of which is to support payments to GPs who claim various monies as of right against the defined *Term of Services,* and to avoid the inflation of GP's list.

II. NHS Central Register databases

The NHSCR established a computerised registration system in 1991 and keeps records on all people known to be alive since 1st January 1991. Historical record systems (manual cards, microfilms and microfiches) go back much further.
remove death from

Each patient's record in the NHSCR database contains the following demographic details:

- Name
- Sex
- Date and place of birth
- NHS number
- FHSA code and date of registration

Figure 2.1 Population registration process within the NHS

- If deceased: date and place of death

Peoples' home address, details of current GPs and clinical information are not recorded in the NHSCR database.

The NHSCR database has been used to assist researchers to integrate patients' demographic data in their local database with that of the NHSCR. Prospective research on patients ('flagging study') has been supported by the NHSCR system. For example, death certificates may be obtained using the 'flagging study'. Details on the 'flagging study' and its costs are described later in more detail. The new NHS number that was introduced in 1995 (where both co-exist) is linked to the old NHS number on the NHSCR database.

2.1.3 Means of data transfer between NHS registration systems

Most General Practices have a microcomputer for keeping demographic data of all their registered patients. Every new registration in these Practices is automatically notified to the local FHSA via the Racal Healthlink. General Practices that are not computerised, will post weekly information on newly registered patients to the corresponding FHSA. The transfer of data between local FHSAs and the NHSCR takes place automatically via the same Racal Healthlink.

Local Registrars deliver weekly notifications of births and deaths to the NHSCR on a floppy disk or on paper format, whilst also sending copies of the death certificate to the NHSCR.

2.1.4 Conclusions

The NHSCR holds the most updated information on the current NHS registration status and NHS number of persons registered with NHS doctors (such as GPs or NHS trust doctors) in England and Wales. Copies of death certificates (the draft entries of all deaths) are received at the NHSCR and the date of death is recorded into its computerised database. In parallel, copies of the death certificates are kept in paper forms and are stored at the Health Authority in Titchfield.

In addition, local FHSAs keep the most updated information on addresses and GP details on all registered patients within their boundaries.

As previously shown, these registration details (patient NHS number, patient current status, address and GP details) are a challenge to the accuracy of data held in the diabetic patients' records in the "Diabeta" database.

It appears that a record linkage of "Diabeta" with the NHSCR will result in the update of all diabetic patients' vital status and new/old NHS number. For deceased patients

the electronic record linkage will also provide details on the date of death. The cause of death will be obtained from copies of the death certificates, also available in ONS.

For the time being, a direct record linkage between "Diabeta" and all the ninety-eight FHSAs across the UK is not cost-effective. Sequential update of the registration details (including patient address and GP details) of all patients records held in "Diabeta" currently cannot to be achieved. Presumably a large proportion of diabetic patients recorded on "Diabeta" are registered with the FHSA in the local area of St Thomas' Hospital: LSL FHSA (Lambeth Southwark and Lewisham FHSA). Thus, a record linkage between "Diabeta" and the local FHSA may result in more complete updating registration details on a proportion of diabetic patients. According to the "Diabeta" database, the proportion of patients currently registered with LSL FHSA is 55%. This estimate was based on a sample of 104 patient records extracted from "Diabeta" database as detailed in the author's MSc project. The percentage of this update may actually be smaller considering the fact that a large proportion of deaths recorded in "Diabeta" failed to be notified in the system database.

In conclusion, a record linkage of "Diabeta" with the NHSCR and a local FHSA will result in a great improvement in the accuracy of demographic data in "Diabeta".

The NHSCR only record patients registered with a general practice in England and Wales. For those who are not registered with a GP (e.g. admitted to hospital through Accident and Emergency), their vital status cannot be traced through the NHSCR. According to the Office of National Statistic's (ONS) reports on vital statistics for population in England and Wales, there were 18.5 million people resident between 1981 and 1998, but only 16.1 million people (87%) registered on the NHSCR. If we include immigrants registered with a general practitioner, the proportion of the real population in England and Wales registered on the NHSCR could be less than 87%. However, most of our patients are referred by a GP and small proportion of people referred by A+E or themselves. Therefore such a link ("Diabeta"-NHSCR) will still be valuable for improving the quality of data recorded on "Diabeta".

2.2 The health care information systems considered for record linkages with "Diabeta"

The health care information systems within the NHS have been continually improved to try to ensure that patients receive the best possible care. In order to know how health services are performed within the NHS and how corresponding information systems support these activities, the functions of relevant health services information systems which are suitable to be linked into "Diabeta", will be illustrated after an introduction to the structure of the NHS.

2.2.1 The structure of the NHS

The current structure of the NHS as illustrated in Figure 2.2 shows that, at local level the NHS is divided into purchasers and providers. The activities of purchasers and providers are overseen by the Department of Health working through the NHS Executive and its eight regional offices. One NHS reform in April 1997 abolished regional health authorities and merged District Health Authorities with Family Health Services Authorities (FHSAs) to form unified Health Authorities (HAs). There are about 60 Health Authorities in the NHS, and these, together with GP fundholders, are the purchasers. They have the responsibility of assessing the health needs of their defined resident population, and buying the health care services needed to meet the health requirements of their local population. Providers are hospitals that still remain under the control of the health authority, or can opt out and become self-governing trusts, together with general practices, communities care, and other service units that supply health care services to those who wish to purchase from them. The pace of change within the NHS is continually increasing, one of the many changes recently being the introduction of Primary Care Groups (PCGs) in April 1999. The aim of PCG is to consolidate GPs, district nurses and other interested organisations (such as Local Councils, Social Services etc.). This is to improve the overall health of their community, make sure that all local people have equal access to all health services, and to plan and buy hospital services which best meet their health needs. PCGs will have the opportunity to become Primary Care Trusts and eventually replace GP fundholders. The details of the organisations "provide & purchase" described in this thesis are those related to the St Thomas' Hospital catchment area - Lambeth, Southwark and Lewisham Health Authority (LSL HA) area.

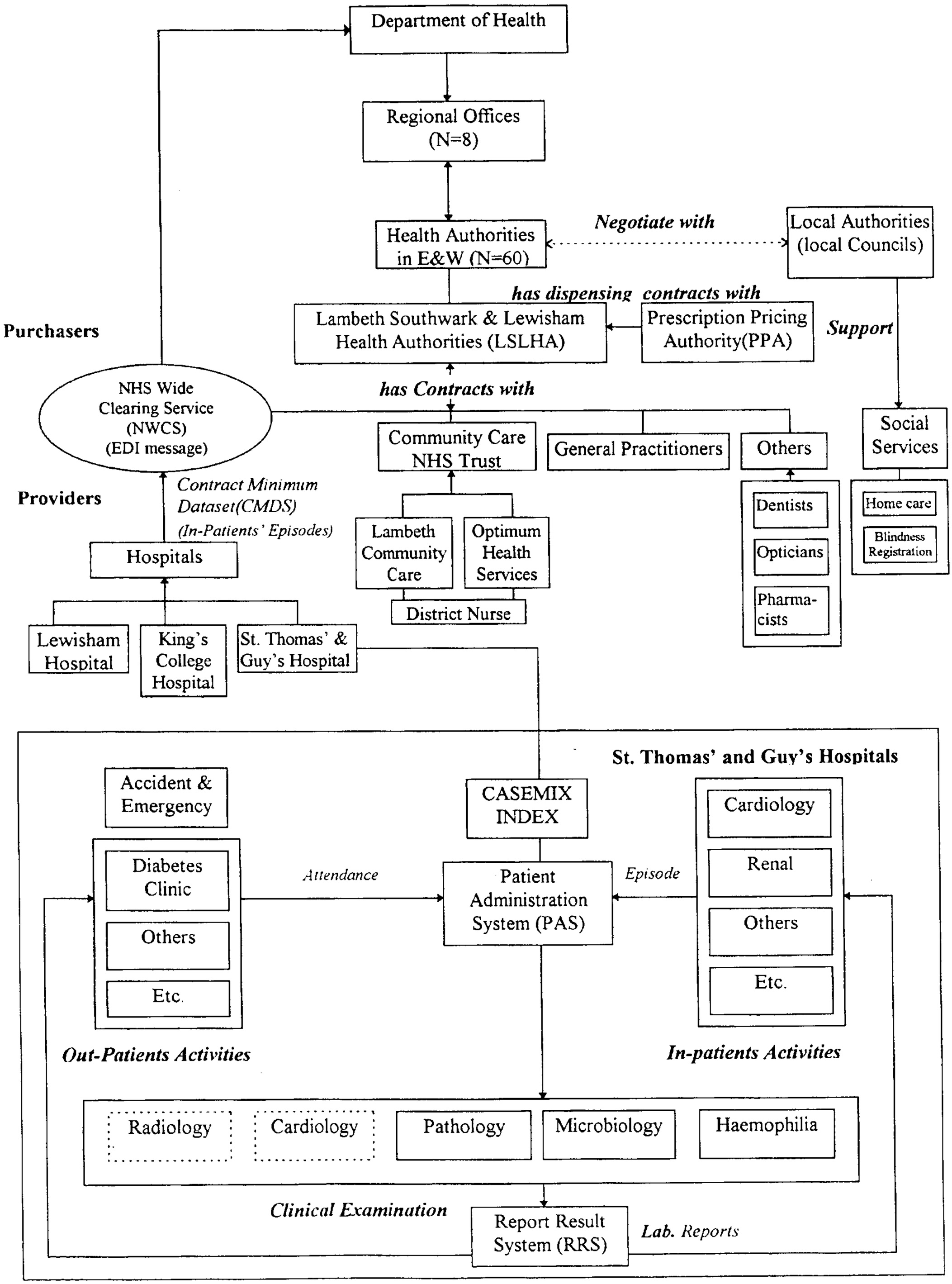


Figure 2.2 Current NHS structure

The activities of health care within the NHS are performed in both secondary (e.g. hospital) and primary care settings (e.g. general practice, community care and other services). Hospital care activities can be mainly divided into three aspects: 1) in-patient activities, 2) out-patient activities and 3) laboratory examinations. General Practices and Community Care (Social Services) provide care/home care as requested for the local residents. The hospitals and other provider units are negotiated via NHS contracts and service agreements with the purchasers, based on the health care activities' data.

Social Services activities, although outside the NHS, are funded by a local authority and mainly provides the domestic care for all kinds of disability (mental or physical (e.g. blindness etc.)), plus learning difficulties. In some circumstances the patient's local council and health authority provide the home care together for the patient, through negotiation.

Since the cutting of health budgets is unpopular, and demand for health services is increasing, successive governments have tried to increase health system activity through administrative reform and resource reallocation, rationing rather than by budget increase. The '1990 NHS and Community Care Act' created a radically different way of providing NHS services. An internal market was born in April 1993, and this introduced the idea of competition to make providers more efficient. As a result, purchasers become more aware of the cost and quality of the services being giving to their population. These revolutionary ideas of internal market, the purchaser-provider split, have made enormous changes occurring in the information required to manage a hospital and in systems designed to deliver that information. The Patient Administration System (PAS) and Case Mix Management System (CMMS) are two systems used in hospitals for useful management. A Contract Minimum Data Set (CMDS) which is generally in the form of PAS/CMMS is held in Health Authorities. CMDS is part of a contracting system which has proved to be effective in health care, however the information they require is different from that needed by providers, who wish to price their service and control their costs. Purchasers, though, may want to compare the relative cost and outcome of competitive treatment. They may also want information on appropriate measures of health gain achieved versus resource

investment, in respect to their populations, so that they can develop and monitor their strategies. The summary of the above information systems and others is described below and also in Figure 2.3 and Appendix A.

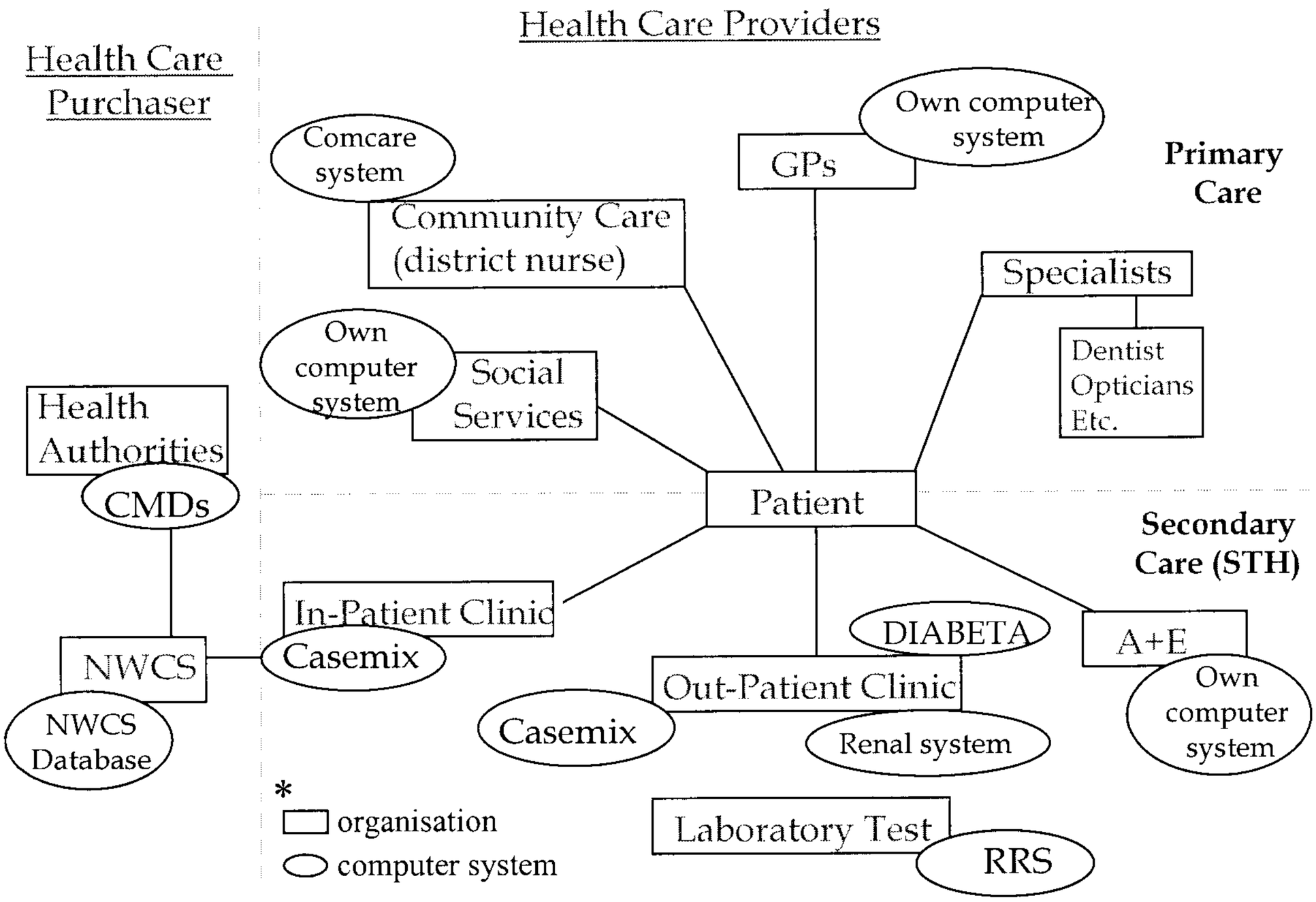


Figure 2.3 Information systems related to the provision of health care provided for patients in the NHS

IT technology is playing an important role in modernising the NHS. The strategy of the New NHS is to put partnership and performance at the forefront of running the NHS, rather than an internal market. This has set out the management and funding mechanisms through which the health strategy has to be achieved. The targets of the NHS have moved on to measure how quality standards will be driven in all parts of the NHS and how efficiency of services can be improved to release resources to further improve quality (The new NHS: Modern • Dependable December 1997). A comprehensive analysis of the information needed to support health services is one of the main objectives of the NHS.

2.2.1.1 Patient Administration System (PAS)

The Patient Administration System (PAS) holds patient-based records involving a wide range of demographic data, hospital activity data (waiting list, out-patient and in-patient or day case information), an array of other useful indicators such as method and source of admission, destination on discharge, method of discharge and contractual data, together with a small amount of clinical information usually in the form of diagnostic and procedure (operation) code. This data is collected into the Patients Administration System (PAS) from wards, out-patient department or consultant's offices. The main functions of the PAS is for booking appointments, sending letters to GP for DNA's (did not attend) patients or generation of discharge summary letters.

PAS usually consists of suites of integrated modules. A patient registration module maintains the person's identification details (e.g. name, date of birth and sex) and holds administrative details (e.g. general practice and home address) of patients treated at the hospital.

St Thomas' Hospital started to use PAS in 1996, as a replacement for its own systems that record in/out-patients activities.

2.2.1.2 Case Mix Management System (CMMS)

In order to 'Improve the quality and quantity of patient care through the better use of resources' (Baird 1992), the Resource Management Initiative (RMI) was launched in 1986 creating an information system that was accessible and credible to all health professionals in the hospital management. At the heart of this information vision lay the Case Mix Management System (CMMS). The data in CMMS is patient-based, so it is possible to obtain a whole-patient picture ranging in theory from basic demographics through operative procedures performed. Most hospital trusts with a CMMS can boast a case mix system with PAS data. Most importantly, these individual patient profiles of longitudinal data can be aggregated to present high-level views of hospital businesses while retaining the ability to 'drill-down' to various sub-levels. It is known that some activities data is not collected into PAS and CMMS (e.g. A+E activities, dialysis etc.), as the large number of visits do not necessarily represent activity, and the payment for these services are relatively financially independent from

hospitals. In this case, the purchasers use their own information system monitoring the health service performance.

CMMS was introduced to St Thomas' Hospital in 1990 and records the hospital in-patient and out-patient's activity at Guy's since 1995 and at St Thomas' since 1997.

2.2.1.3 Contract Minimum Data Set (CMDS)

With the advent of the internal market, contracting systems are becoming increasingly important to NHS managers. The Contract Minimum Data Set (CMDS) is a subset of CMMS, which excludes some local details (such as specialist code). The data set submitted from the hospital (provider) to the corresponding HA (purchaser) in which area patients live, is via a 'Local Clearing Services' (LCS) system under control of the Regional Health Services Information Management Group. This LCS is not only responsible for routing the CMDS from provider to the corrected HA, but also downloads the CMDS to a database recording all hospital in-patient episodes occurring within its region. Each LCS also sends CMDS to the Department of Health (DoH) for hospital episodes statistical purposes. Since April 1997, the NHS-Wide Clearing Services (NWCS) have replaced these LCSs, and downloads all hospital in-patient episodes supplied by each health authority (in England and Wales) to the NWCS database held in the DoH, attaching a new NHS number to each record. Therefore the DoH now holds CMDS for all patients in England and Wales. (Figure2.4)

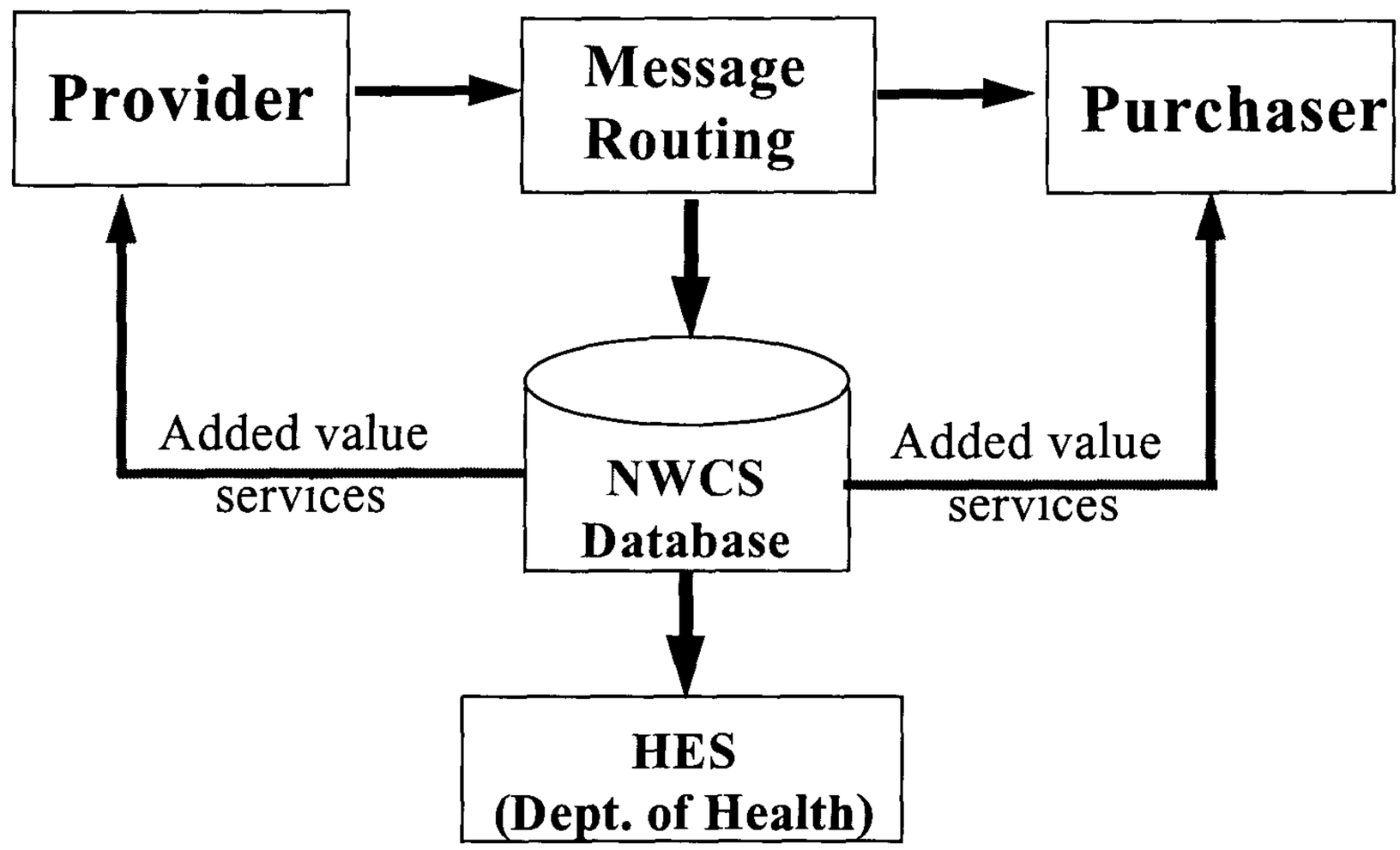


Figure 2.4 Protocol of NHS-Wide Clearing Services

Lambeth Southwark and Lewisham Health Authority is in the catchment area of St Thomas' Hospital. The CMDS in LSL started recording episodes' data in 1987.

It can be summarised that the resources related to the hospital in-patient activity are recorded in CMMS (hospital based records system) and CMDS (district, region and national based database). The linkage of "Diabeta" to these systems will be discussed in the next section.

If "Diabeta" is linked with CMMS & CMDS, information on the number of days spent in a hospital bed, the particular treatment under a specialist (consultant code, procedures code), and treatment related to a complication of diabetes can be obtained on "Diabeta".

2.2.1.4 Accident and Emergency information system

Accident and Emergency (A+E) in a hospital setting has an independent contract with the local Health Authority. It has its own computer system to record activities performed in it's own A+E. The diagnosis is not recorded in the records system although a free text record of the problem is available.

The computer system in A+E at St Thomas' Hospital records all A+E attendance since 1992.

2.2.1.5 Out-patient information system

PAS also collects records of all out-patients' attendance for hospital management purposes, and this is downloaded to CMMS for audit. However, the information on dialysis is not normally collected into PAS or CMMS, because the accumulated numbers of dialysis sessions cannot be used to measure the outcome of care. Its own system ("Proton") has been developed to obtain the details of renal treatment.

St Thomas' Hospital used its own system – COPS - to record hospital out-patients' treatment's activities since 1987, and in 1996 the PAS has replaced these two systems and has transferred records to the CMMS since 1997.

2.2.1.6 Laboratory reports information system

The Results Reporting System (RRS) is one of the hospital information support systems which collects current laboratory reports from microbiology, pathology, chemical pathology, radiology, haematology and haemophilia departments. The Patient Master Index (which is from the PAS module) allows all tests related to one individual to be cumulatively recorded.

St Thomas' Hospital started to use the RRS system in 1995, but originally it only kept the computerised records for six months; it is now able to keep them for one year. radiology and cardiology reports in St Thomas' Hospital will be soon collected into the RRS as well.

2.2.1.7 General Practice information systems

As mentioned earlier, the GP is responsible for providing treatment for local residents within the practice and also for referring patients to a hospital or communities care service where necessary. As the FHSAs remunerate GPs against a centrally negotiated fee-scale reflecting the agreed contract (Department of Health 1989c), the FHSAs predominately allocate funds from a budget which is not cash limited, against which GPs claim various payments as of the right defined in "*Terms of Services*". Therefore it is necessary for the GP only to show FHSAs that a service has been provided. The age, gender and address of each patient is held in the database so that capitation payments may be made to the GP, and likely demands on general medical services as a consequence of the demographics of a particular list, can be deduced.

In order to achieve accuracy and efficiency in the process of contracting, General Practices have developed computerisation to a higher degree in comparison to FHSA's. Almost 85 per cent of General Practices have, on average, more than two computers (NHS handbook 1999/2000).

Computers usage in general practice began with a few enthusiasts, but as early as 1983, the 'Micros for GPs' scheme encouraged practices to obtain machines to establish their potential. Since then, various GP computer systems (e.g. VAMP, AAH MEDITEL and EMIS etc.) have been developed as part of the NHS reforms in purchasing, medical audit, preventive care and health promotion. These systems record demographic data, GPs activities data (such as visits to GP), referral status and letters received from other health care providers (hospital, community care etc.) and repeat-prescription data. These systems have been developed from captured demographic data only, to clinical data, during the consultation. By 1993, over 90 per cent of practices kept clinical data, allowing three-quarters of practices to use their computers for audit and one-third for research (NHSME 1993b). At present a number of diagnostic classifications are in use including the Read code system and the Royal College of General Practitioners (RCGP) codes etc. This enables GPs to code the primary care diagnostic and all activities performed in the practice. In this case, it is possible to identify if the visit is related to diabetes from such a system. However the poor data quality as a result of inadequate training and lack of motivation on the part

of the GPs occurred in the early stage of the development. Although computer systems are widely used by GPs, paper records still remain largely in use and in parallel.

In conclusion, a records linkage between "Diabeta" and GP systems could ascertain the number of visits to GP, prescription of medications and out-patients attendance in hospital/community care services on diabetic patients registered on "Diabeta". This is particularly important for outcome studies.

2.2.1.8 The information system used in Community Care within the NHS

Community health services encompass a complex set of services. They are managed by a 'web' of NHS trusts under a contract with the local health authority, mainly providing home medical care for disabled and elderly local residents. The core elements of community care are available from district nurses and health visitors. Other service provisions are by arrangement.

For contracting purposes, the information system has been recently updated and is used in most community care trusts. It is called the "Comcare" system which only records the type of visit and the services provided to local residents in free text. Therefore the services provided by communities care can be only accounted for by the number of contacts with a district nurse. Within the St Thomas' Hospital catchment area, there are three community care trusts, but only one of them, Lambeth Health Care, started to use "Comcare" to record those home visits provided by district nurses for residents living within the Lambeth area in 1995. If "Diabeta" is linked with "Comcare" in Lambeth Health Care, the number of home visits by district nurses for our patients from the Lambeth area can be defined in "Diabeta".

2.2.1.9 Local Authority Social Services information system

Social care is funded by a local authority providing mainly domestic care for all kinds of disability (mental or physical (e.g. blindness etc.)), and also learning difficulties. The information system used in each Social Services Department records the social care activities provided to local residents. With regards to the complications of diabetes, the registration of blindness in a Social Services data set might well be useful. Although the completeness of this registration cannot be one hundred percent

(because of the registration of blindness depending on patients wanting disability benefits from Social Services), it may be the only resource to ascertain the indirect cost of the diabetes complication of blindness.

The Social Services in Lewisham Council has the most complete computerised records system within LSL HA area, and have, in conjunction with manual records since 1990, identified all people who live within the Lewisham area requiring a Social Service. If "Diabeta" is linked with the computerised records system in the Lewisham Council, the number of attendance to Social Services and the number of registered blind patients in the Lewisham area can be ascertained.

2.2.2 Means of data transfer between health care information systems within the NHS

The health authorities, as health care purchasers, collect contractual information (e.g. hospital episodes, item services and pathology report etc.) from health care providers such as hospitals, general practices and community care settings etc. through a local network - Racal Healthlink. The Contract Minimum Data Set (CMDS) is originally collected from a Patient Administration System (PAS) via a hospital network through each individual clinical department. The data set from PAS is downloaded into a hospital Case Mix Management System (via hospital network). It will then be sent to the NHS Wide Clearing Services (NWCS) system so as to convert the different format data into a standardised electronic data interchange (EDI) message and then directing them to reach the corresponding health authority (via Racal Healthlink). The hospital laboratory reports (e.g. microbiology, pathology, chemical pathology, radiology, haematology and haemophilia) are transferred from each laboratory to the Results Report System (RRS) via the hospital network continually until they become available. In some hospitals, the pathology reports have also been transferred to corresponding GPs system via the local network (Racal Healthlink).

2.2.3 Conclusions

Various information systems are working together to support the performance of health care provided in the NHS. CMMS, CMDS, and NWCS database record a minimum data set of hospital in-patient activities (known as a hospital episode) and some clinical data, which is a subset of PAS data in each hospital. In addition, the hospital systems such as CMMS, RRS, Accident and Emergency system, "Diabeta" and renal system ("Proton") record all out-patients' clinical activity data (known as out-patients attendance data) in a hospital. Other local systems such as GP systems, "Comcare", and Social Services systems record the community care activities for local patients.

As previously explained, the diabetes care activities data (such as hospitalisation, attendance in Accident and Emergency and community care etc.) are a challenge to the completeness of data held in the diabetic patients records on the "Diabeta" database.

It appears that a record linkage between "Diabeta" and the NWCS in the DoH will result in the collection of all diabetic patients' hospital in-patient episode data, but only for those events occurring after 1997. Other systems (e.g. CMDS, CMMS, A+E system, RRS, GP, "Comcare" and Social Services system) record the clinical activity data (in/out-patient clinical activities, accident & emergency, laboratory test and community care), but only for the local population. The range of population recorded in the computerised system and the time the computerised system was introduced has become a restriction for consideration of record linkage for "Diabeta". Fifty-six percent of those patients registered with St Thomas' Hospital are local residents, and more likely to attend local hospitals (such as at St Thomas', Guy's, Kings College and Lewisham Hospital) and community care institutions. It is probably an unrealistic idea to link "Diabeta" with all health authorities or hospitals and communities care systems in England and Wales. Linking "Diabeta" to the following systems, however, would at least allow us to access the activities' data for those patients living within the St Thomas' catchment area.

1. CMMS in St Thomas and Guy's Hospital
2. Contract Minimum Data Set (CMDS) in Lambeth, Southwark and Lewisham Health Authority
3. Accident and Emergency in St Thomas' Hospital
4. Renal system ("Proton") in St Thomas' and Guy's Hospital
5. Result Report System (RRS) in St Thomas' Hospital
6. "Comcare" in Lambeth Health Care
7. Social Services system in Lewisham Council

Appendix A shows that the clinical data items that are recorded in these information systems, and summarised in Table 2.1.

Table 2.1 Clinical data recorded in all relevant information systems

Services	**IT system**	**Diagnosis Recorded**	**Treatment recorded**
Hospital in-patients activities in St Thomas and Guy's Hospital	CMMS	yes	yes
Hospital in-patients activities in Lambeth Southwark and Lewisham Health Authority	CMDS	yes	yes
Accident + Emergency in St Thomas' Hospital	own computer system	no	no
Dialysis in St Thomas and Guy's Hospital	"Proton"	yes	no
Laboratory reports in St Thomas' and Guy's Hospital	RRS	no	no
Care in General Practices in UK	own computer system	yes	yes
Home care provide by district nurses in the Lambeth area	"Comcare"	no	yes
Social Services in the Lewisham area	own computer system	no	yes
Diabetic Clinic in St Thomas' Hospital	"Diabeta"	yes	yes

If "Diabeta" were to be linked with all above relevant information systems, the encountered events (number of re/admissions, visits to GP and A+E in STH and Lab. tests in STH), drug dispensing and specialist treatment (e.g. cardiology, vascular, ophthalmology and chiropody) can be ascertained.

The technical possibility and cost incurred for linking "Diabeta" with these relevant systems will be discussed in Chapter 3.

Chapter 3

Evaluation of the feasibility of record linkages between "Diabeta" & other information systems

The aim of linking the records in the "Diabeta" database with the validated data held in other information systems is to obtain, as much as possible, accurate and complete data recorded on "Diabeta". This will enable "Diabeta" to *update* the demographic data that is susceptible to change over time, and to ascertain all health care activities and clinical data occurring outside the DEDC.

Before any conclusions are given on the feasibility of record linkages between "Diabeta" and the relevant information systems, some general theoretical aspects on records linkage will be briefly explained and also considerations on the type of the linkage to be chosen in the specific case of "Diabeta".

3.1 The concept of record linkage

3. 1. 1 Means of data transfer via a records linkage

There are several methods of data transfer via a record linkage between two databases, classified as indirect and direct.

1. Indirect linkage

- *Paper*

  Data filled out in the format of hand written or typed media is transferred between organisations by post.

2. Direct linkage

- *Magnetic medium* (floppy disk, tape)

  Data formatted on to a floppy disk/tape that is ready for computer data processing. The data communication between the two databases is undertaken by sending a floppy disk/tape by post.

- *Electronic Data Interchange (EDI)*

  Data formatted as an electronic message in an agreed format is communicated by local/national network (LAN/WAN).

- *Telecommunication*

  Data formatted as an electronic message is downloaded to a telephone line, thereby achieving data communication between remote systems (internet).

The type of record linkage applied for transference of data depends on the comparability of the two computer systems and in which format the data is supplied.

3. 1. 2 Updating records via a linkage between two databases

Any records updated via an electronic linkage between two information systems relies on a process of transferring information from a **'transaction file'** into a file (**'master file'**) that needs to be updated. The product of the update process is a **'new master file'** or **'son file'** and is processed by a specific **'update program'**. Prior to the actual update, data that in **'transaction file'** must be validated by a **'validation program'**.

In the case of this study, the records in "Diabeta" are represented as a **'master file'** that needs to be updated by records identified as **'transaction files'**, from the relevant information systems. The updated "Diabeta" records are defined as the **'new master file'** or **'son file'**.

The **'update program'** relies on matching validated records in the **'transaction file'**, with those records in the **'master file'**. The aim of matching these two sets of records is to bring together information from two resources that apply to the same individual.

In order to achieve this, information that uniquely identifies the persons in both sets of records must be established, e.g. NHS number, patient name, date of birth etc. The data set needed for matching is normally based on the person identification data held in the **'transaction file'**. Therefore each incoming record provided from the **'master file'** will be searched by the **'update program'** for the identifiers in the **'transaction file'** and subsequently these identifiers will be searched in the **'master file'**. Finding this set of identifiers in the **'transaction file'** corresponds to a 'successful match'.

In practice, a group of identifiers may not be entirely found in the two sets of records (e.g. only patient name and year of birth were found common in the two sets of records (**'transaction file'** and **'master file'**)). The rest of identifiers are either missing in a record provided by the **'master file'** ("Diabeta") or are in disagreement with their analogues in the **'transaction file'**.

It is unlikely that all the required patient identifiers in the **'transaction file'** will be available in the **'master file'** for record matching performance. Therefore the amount of details that can be supplied and the proportion of the study population for whom the entire identifying data set is available have an important role on what technique of record matching can be used.

The matching techniques are classified into three categories: i. **automatic matching**: record matching performed on a computerised record system, automatically updated, ii. **semi-automatic matching**: record matching performed on a computerised record system with an operator interfering iii. **manual record matching**: record matching on a historical record system (e.g. paper file and index card etc.). The type of record matching technique used depends on the quality and quantity of identifiers provided. The unique identifier (e.g. new NHS number) is a key to the advancement of the record matching process. In the event of no such number, tedious and cost-incurring techniques are subsequently used, often with poor results.

The general update of demographic data, health care activity data and some clinical data in "Diabeta" records is discussed below.

3.1.2.1 Updating demographic data in the "Diabeta" records by the NHSCR records in ONS

The **'transaction file'** is represented by a record held in the NHSCR and is used for updating data in its corresponding record[1] from "Diabeta".

The **'master file'** is represented by a record extracted from "Diabeta" and is updated in ONS. The update of the 'master file' ("Diabeta" records) by data from its corresponding 'transaction file' (validated NHSCR records) is carried out by a special update program

available in ONS. This update program relies mostly on the ONS computerised system, CHRIS (Central Health Register Information System) which holds the central index of NHSCR. For cases that cannot be traced on CHRIS, the historical NHSCR records system will be used (Figure 3.1).

The '**new master file**' or '**son file**' is represented by the 'updated "Diabeta" record'.

The process of updating a 'master file' ("Diabeta" records) in ONS consists of the following three steps:

1. Amending the existing data in the original file.
2. Inserting new data (e.g. an old & new NHS number)
3. Deleting data no longer required (e.g. if "Diabeta" has the codes of FHSAs with which patients are currently registered, such codes would be deleted for patients found to be deceased).

The **'update program'** on the NHSCR provides the possibility of linking records in "Diabeta" with the corresponding records in the NHSCR. The number of successful records matched will depend directly on the quality of patient identifying data provided to the NHSCR by researchers.

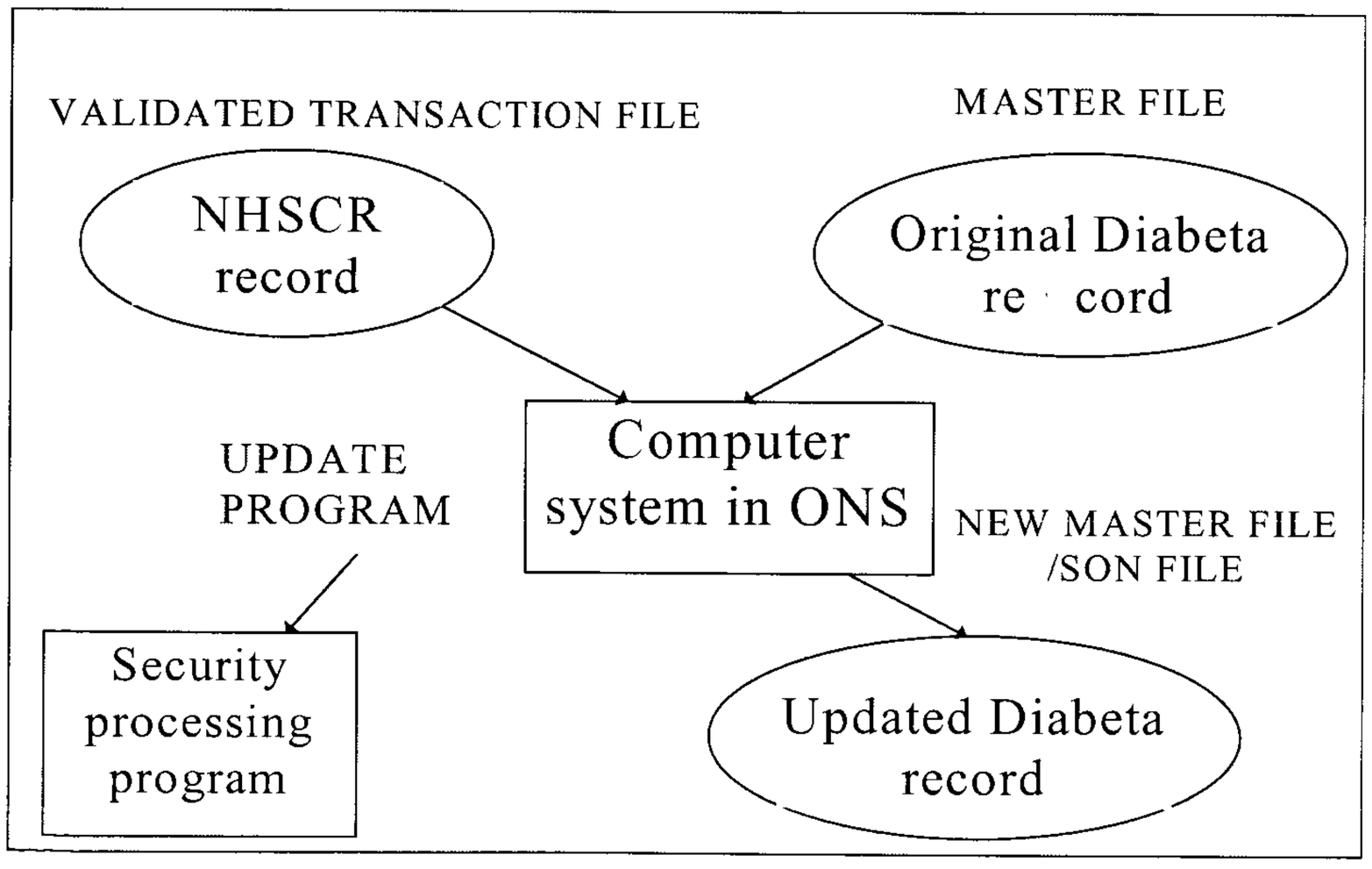


Figure 3.1. General records update process (applied to the particular updating in ONS)

[1]Corresponding records refers to the same person whose records recorded on "Diabeta" and also found on the NHSCR database

3.1.2.2 Inserting health care activity data occurring outside the DEDC in "Diabeta" records by using other health services information systems (CMDS, CMMS, A+E, RRS, GP's computer systems etc.).

The updating process ascertaining to health care activities and clinical data occurring outside the DEDC in "Diabeta" records by other relevant information systems, is similar, as it applies to demographic data in "Diabeta" records performed by the NHSCR records in ONS.

In the updating process, instead of using the NHSCR records as a '**transaction file'**, the validated records held in the relevant health services information systems (such as CMDS, CMMS, A+E, RRS, GP's computer systems etc.) are used to insert the health care activities occurring outside the DEDC for corresponding patients recorded in "Diabeta". The records extracted from "Diabeta" which are to be updated by the relevant health services information systems are still represented as '**master file'**. The update process is carried out by an **'update program'** using a process of matching validated records held in the relevant health services information systems with those provided by "Diabeta".

Updating a **'master file'** in "Diabeta" will insert the activity data (e.g. the date of attendance/ admission and discharge etc.) in "Diabeta", and also update the "Diabeta" records (e.g. diagnosis or operation procedures) given by health professionals outside the DEDC.

3.2 Matching records between two databases

The review of the theory of record linkages illustrates the process of updating the "Diabeta" records by the relevant information systems. In order to know if the linkage is technically feasible and cost-effective, the following questions need to be answered:

1. To what extent can patient identifiers be used in the record linkage?
2. What format of linkage can be obtained (direct, indirect)?
3. What is the cost of linkage compared to its benefits?
4. When does 'data confidentiality' apply?

3.2.1 Means of matching records in ONS

Ideally, the data set requirements for tracing individuals on the NHSCR are:

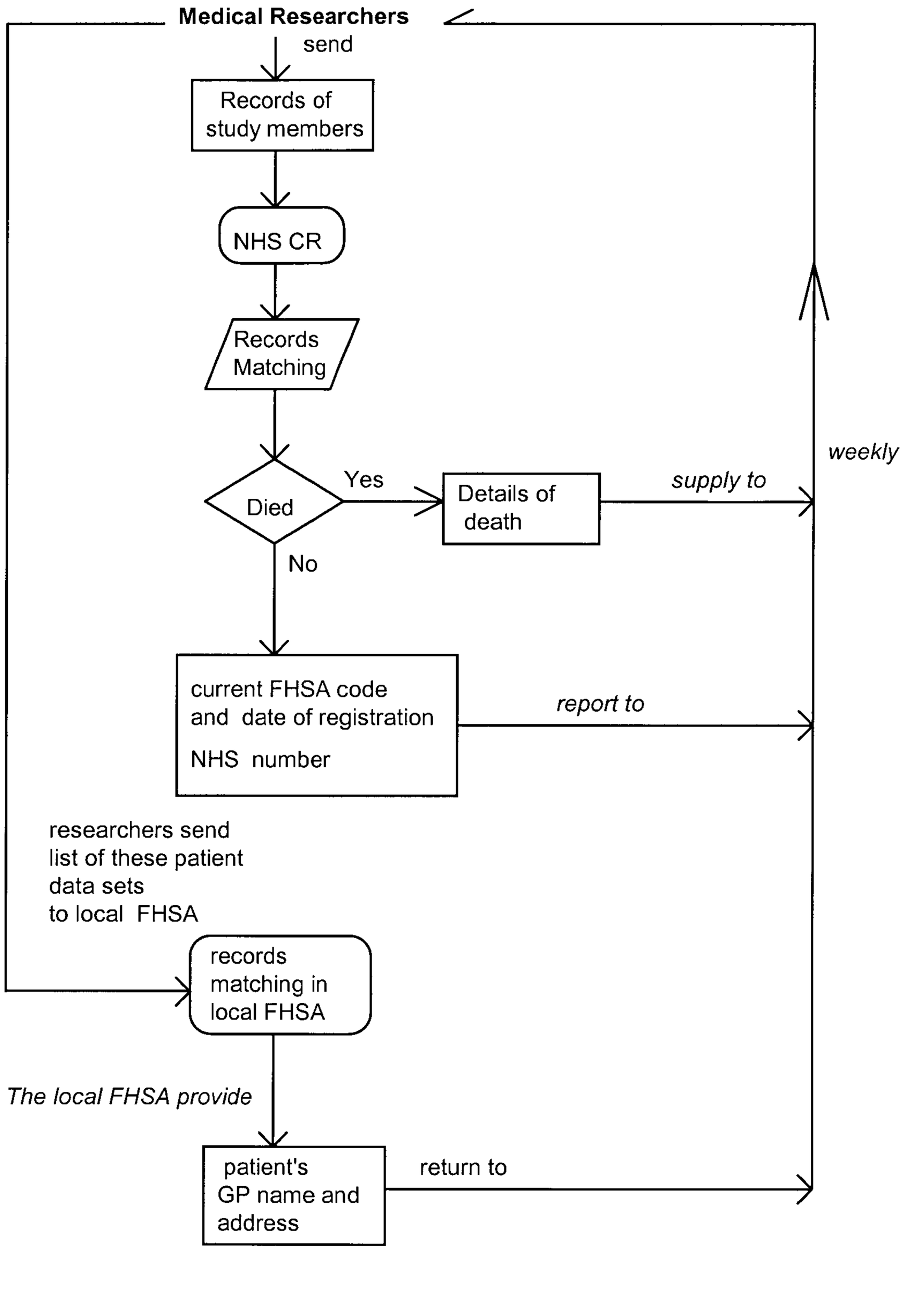
- NHS number
- Full name (not initials) and surname
- Forename
- Maiden name or former surname of a woman who has been married, and title i.e. Miss or Mrs.
- Exact date of birth (essential where NHS number is not provided)
- Place of birth
- Last known address and dates

Provision of all of these data items will result in a positive identification (successful match).

Since it is unlikely that all the required patient identifiers will be available for every subject in any one particular study, requests for an update should not be rejected on these grounds alone. However, the amount of detail that can be supplied and the proportion of the study population for whom the entire data set is available, still has an important role on the techniques used by ONS to perform that study. It is important that any request for update should state the number of cases to be traced and the identification particulars available to avoid unnecessary and tedious cross-matching process. After a successful record matching achieved on the NHSCR database, a 'tracing' and 'flagging' study can be performed. This is to update records supplied by researchers both retrospectively and prospectively.

a) **Tracing studies**

The purpose of these studies is to trace in the NHSCR, records of patients who might otherwise be lost to follow-up. Tracing studies result in the identification of the FHSA in which the patient is currently registered with a doctor. Researchers can then write directly to the relevant FHSA to enquire about the GP who is responsible for a particular patient. The administrator of that FHSA will identify that GP. The FHSA administrator will, however, invariably refuse a request for the address of the patients, as stated by statutory requirements (Figure 3.2).



* study member is refered to as the patient whose record is searched

Figure 3.2 'Life cycle' of 'tracing' study.

b) **Flagging studies**

The purpose of the 'flagging' studies is to prospectively identify the subjects of the study population as soon as these subjects become available for the purpose of particular research. 'Flagging' studies are performed in ONS as follows:

The NHSCR records of the members in the study population are 'flagged' with a code in that corresponds with the purpose of the specific research study. ONS will inform the researchers involved, as soon as the event relevant to the purpose of that study is notified in the 'flagged' records.

For instance if the "Diabeta" database is subjected to a 'flagging' study to update the vital status of patients in its database, all the records in the NHSCR identified as belonging to patients recorded in "Diabeta" will be flagged with a specific code. This code will encrypt the purpose of the study (e.g. updating vital status). When deaths are notified to the NHSCR and are found to relate to patients in the "Diabeta" database, the researchers from St Thomas' Hospital will be informed of the death and supplied with details of death (death certificate).

Researchers can also be informed of other events that may occur after 'flagging', e.g. whether the individual is currently registered with a NHS doctor, and if not whether this is because he or she has left the country or for other reasons (Figure 3.3).

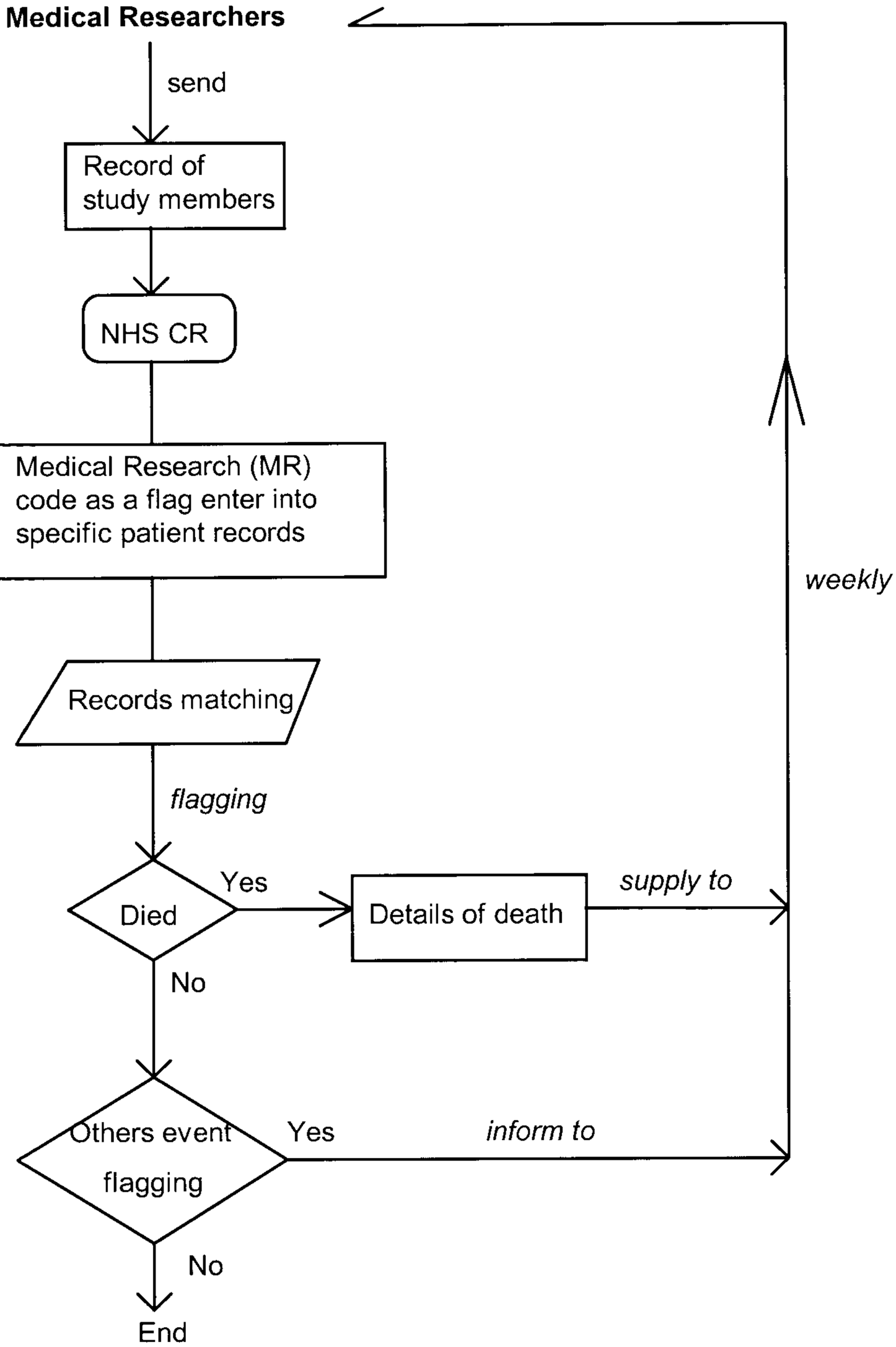


Figure 3.3 'Life cycle' of 'flagging' study in the NHSCR

3.2.2 Means of matching records in the hospital information systems

The patients' identification data recorded in the relevant health service information systems is described in Table 3.1. A unique identification number is a key in the record-matching process. It has shown that a local patient identification number (hospital number) is available in all the information systems within St Thomas' Hospital. The hospital number is issued to all new patients who attended their first appointment in a hospital clinic by the Patient's Master Index system, which gives only one number for each patient. This has enabled us to match records between "Diabeta" and patient records systems at St Thomas' hospital (e.g. CSSM, RRS, A+E own computer system, and renal system ("Proton")) automatically based on the hospital number. A hospital local network which supports the data transference between the internal systems is currently available for transferring updated data between systems within St Thomas and Guy's Hospital (e.g. between "Diabeta" and CSSM or RRS), with the exception of A+E computer system and Renal system ("Proton").

Table 3.1 Identification data in hospital information systems used for record matching

System name	**Data used for matching Records**	**Type of Matching**
Accident + Emergency (STH)	• Hospital number • Patients name • Date of birth • Address	Manual
CMMS (STH & GUY)	• Hospital number	Automatic matching through hospital network
Result Reporting System (RRS) (STH & GUY)	• Hospital number	Automatic matching through hospital network
Renal system ("Proton") (STH & GUY)	• Hospital number • Patient's name • Date of birth • Address	Manually matched

We conclude that the accuracy of record matching between "Diabeta" and other information systems in Guy's and St Thomas' Hospital is greater than information systems outside of the hospitals, and more advanced matching techniques (automatic matching) can be applied, when based on a unique patient number.

3.2.3 Means of matching records in an Health Authority system

The identification data held in the CMDS of a health authority is listed in Table 3.2. Patients' names and addresses (but not post code) are excluded in data set. Moreover, a unique patient number is unavailable in the CMDS (It can normally be used for matching data from CMDS with the corresponding records from a local database (e.g. "Diabeta")). Nevertheless, the unique patient number (the new NHS number) has only been recorded on CMDS since April of 1997. Therefore the records matching process relies mainly on other patient identification data.

The postcode recorded in CMDS is related to the address at that time of the patients' admission. Nevertheless, that fact that "Diabeta" keeps patients current and previous address may increase the probability of successful records matching on the CMDS historical records system. In this case, the completeness of a current or previous postcode record on "Diabeta", will be a crucial issue.

In theory, it is now possible to match all our diabetic patients' records with CMDS and NWCS database in the department of health based on local patients' identifiers/ new NHS number, and to obtain information on in-patients activity on our diabetic patients.

Table 3.2 Identification data in health authority information systems used for record matching

System name	**Data used for matching Records**	**Type of Matching**
Contract Minimum Data Set (CMDS) (LSL HA)	• Postcode • Sex • Date of birth The new NHS number (only available for records since April 1997)	Automatic matching. Transfer data could be an EDI format through NHSnet in future

3.2.4 Means of matching records in community care systems

A new NHS number has been generated into all GP's computer systems since 1995. Every patient who is registered with a GP has a new NHS number assigned by the HA or NHSCR. If "Diabeta" can obtain the new NHS number from the NHSCR or hospital Patient Master Index, the record matching between "Diabeta" and GP systems will be mainly based on the new NHS number. The practice code in "Diabeta" can be used to group patients into the same practice, and then the patient's identifiers (NHS number or patient's name and date of birth etc.) from "Diabeta" can be sent via a local network or magnetic medium to the corresponding practice for records matching. For those local practices which have no computer facilities, or those who have a small number of patients registered this may be best achieved as an indirect linkage (such as sending data collection forms), and may prove to be a better solution. The accuracy of recording patients' current GP on "Diabeta" will be crucial in achieving a successful linkage. Nevertheless, the current patient's GP details can also be ascertained through the linkage between "Diabeta" and the NHSCR.

The records matching between "Diabeta" and "Comcare" / Social Services system can be based on the patient's name, date of birth, address and GP name etc. through a magnetic medium. Patients' records can be selected via a Geographical Information System (GIS) and sent to the corresponding local organisations for records matching.

Table 3.3 Identification data in community care information systems used for record matching

System name	**Data used for matching Records**	**Type of Matching**
General Practitioners (England & Wales)	• New NHS number • Patient's name • Address • Date of birth • Practice code (or postcode) • GP details	Sending a letter to each GP with requirements of patient's data.
"Comcare" (Lambeth Health Care)	• Patient's name • Date of birth • Address	Manual matching Data transfer by a floppy disk
Social Services (Lewisham Council)	• Patient's name • Date of birth • Address	Automatic matching Data transfer by floppy disks.

3.3.1 The cost of record linkages

The cost of record linkages is mainly estimated according to which records matching technique is used in the matching process.

3.3.1.1 The cost of medical research in ONS

As previously explained, ONS uses different procedures for matching records which are suitable for 'tracing' and 'flagging' studies according to the quality and quantity of patient identifying data provided by researchers.

ONS has four categories of charges (bands of costs) depending on the difficulty of the matching process and the matching procedures involved.

Band A: For records automatically matched on the NHSCR computer system, e.g. from disk without operator intervention

Band B: For records matched by operators on the computer system.

Band C: For records not matched on the computer system but bearing a valid NHS number.

Band D: For records not matched on the computer system and not bearing a valid NHS number.

The charges in the bands of costs *C* and *D* are for those records identified by the NHSCR historical records.

The prices for each band are displayed in the table below.

Band	**price** (per study member)
A	0.30
B	1.70
C	3.70
D	4.75

Cases require 'flagging' for prospective notification of death, cancer and miscellaneous events as soon as they occur.

It appears that the more advanced the matching technique (e.g. automatic matching) applied, the less the cost of linkage obtained. If "Diabeta" can provide better quality of identification and 'flagging' for patients alive since 1992 (year of introduction of computerised records in NHSCR), the linkage between "Diabeta" and the NHSCR will become cheaper.

3.3.1.2 The cost of records matching in other information systems

The cost of these record linkages is not estimated, owing to the fact that "the price" of records matching is not defined by each system. In general, the cost of record linkages within St Thomas' Hospital will be cheaper based on a hospital number. Conversely, it will be more expensive unless a unique patient number (such as new NHS number) is applied. However, as mentioned earlier, most of our patients are from a local area and are more likely to have hospitalisation in a local hospital (e.g. St Thomas' or Guy's Hospital), if required. Even for those patients who moved outside the St Thomas' Hospital catchment area, most of them still attend services in their original local health service settings. Records matching between "Diabeta" and a local system (e.g. GPs and HA within LSL) may be relatively easier and cheaper. A cost-effectiveness study on a local diabetic population is required to confirm this possibility.

3.3.2 Data confidentiality

Patient's demographic data linked with clinical data are very sensitive. The security and confidentiality of patient records is extremely important when linkages between "Diabeta" and NHS information systems are implemented.

Handling patient's data must comply with the principles of the Data Protection Act. Based on this, each NHS information system collaborates with other patient data holders by defined standards, and by procedure manuals on data confidentiality and security.

In the case of records linkages between "Diabeta" and the NHS information systems, the NHS will state the measures which must be taken by that system's owner to ensure

that only authorised and agreed data are exchanged. The patient's consent needs to be obtained for accessing clinical data held in General Practice, Community Care and Social Service settings.

Special sets of rules are applied for other different means of data transfer in order to protect the patient's personal data from disclosure to a third party.

When data transfer is in the form of a magnetic medium (e.g. floppy disk or tape), the symmetric technology should be used to arrange a secret key for the authoriser, who alone, is allowed to read the floppy disk (data security processing).

3.3.3. Frequency of data transfer between "Diabeta" and other NHS information systems

The frequency of data transfer between "Diabeta" and information systems within the hospital can be transferred on a daily or weekly basis via the hospital network based on patient's hospital number. But to link with information systems outside the hospital, it will probably be a month, because of various transformation formats (e.g. local network, NHS network, floppy disk or paper format etc.) used within NHS information systems and where no unique patient's identification number has been fully used yet. The frequency of data transfer between "Diabeta" and other NHS information systems can be technically increased when the new NHS number is available in all systems.

Section Two

Methodology

In this section, the feasibility protocol of linking "Diabeta" with the NHSCR and other information systems within the NHS, based on two cohorts of patients, will be described. The methodology used for 'flagging' a cohort of patients on the NHSCR and tracking a cohort of patients in other information systems will be demonstrated. In order to illustrate the benefits of the linkage ("Diabeta" - NHSCR) in respect to epidemiology study and medical research, certain statistical studies and a Geographic Information System (GIS) techniques were applied. The functions of which will be explained in this section.

Chapter 4

Research Protocol

A pilot study on the feasibility of linking "Diabeta" with the NHSCR has been evaluated in the author's previous study (MSc project), based on a small cohort of diabetic patients (n=104). The pilot study recommended that a linkage between the NHSCR and the entire "Diabeta" database can and, should be implemented. The benefits of this link in terms of epidemiology and medical research are evaluated in the present study. In order to do so, the following research activities were performed:

1. Over 7000 records (1979-1994) were extracted from "Diabeta" and formatted on to floppy disks (as suggested by the NHSCR researchers) and sent to the NHSCR.
2. Design of a database containing fields on patients' demographic and death certificate details written in Microsoft Access (Figure 4.1).
3. Transference of identification data of the 7000 patients from "Diabeta" to the database.
4. Entering the updated demographic data and death certificates of the 7000 records received from the NHSCR onto the database, by matching patient's hospital numbers.
5. Analysis of the results of the records linkage between "Diabeta" and the NHSCR.
6. Analysis of the benefits of this record linkage in terms of epidemiology studies.
7. Ways of solving existing errors in the records linkage ("Diabeta" - NHSCR).
8. Evaluating the sensitivity and specificity of the linkage ("Diabeta" - NHSCR).
9. Strategies of linking "Diabeta" with a local Health Authority patient registration database.
10. Investigation of the development of the NHS wide network.
    This was to investigate the possibility of moving the existing linkage ("Diabeta" - NHSCR) to the NHS wide network in the future.
11. Using a geographic computer software (MapInfo) to establish relationships between demographic location and morbidity & mortality in the diabetic population.
12. Recommendations about the further development of "Diabeta".

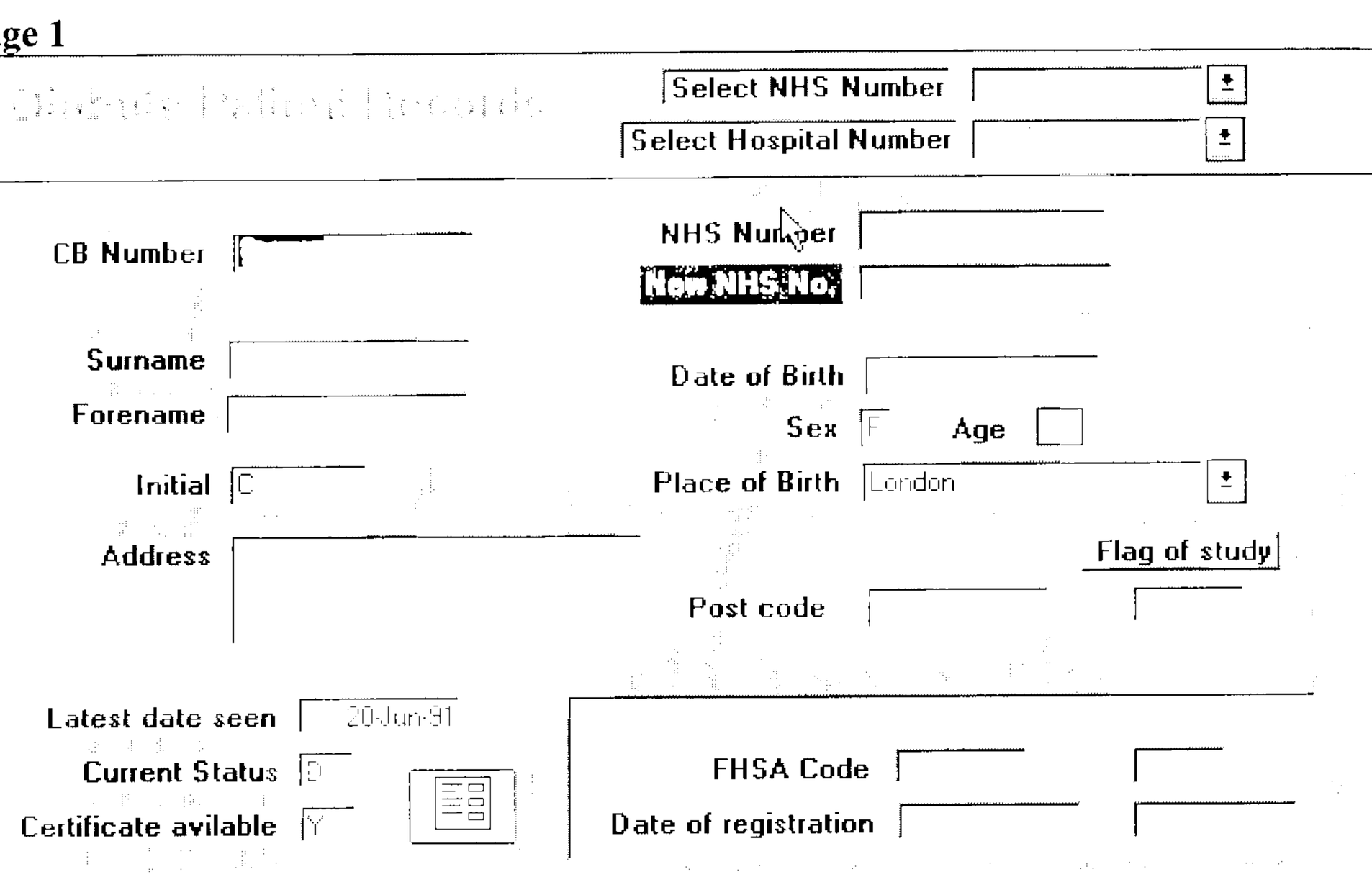


Figure 4.1 A Screen of the 'Microsoft Access' database used for entering the updated data from the death certificate received from the NHSCR.

In order to further improve the accuracy and completeness of patient data held in "Diabeta", the author investigated the possibility of linkages between "Diabeta" and other information systems within the NHS.

1. Investigation of patients' records systems within St Thomas' Hospital, such as CMMS, Accident & Emergency, laboratory report system etc. This was to determine if they can be suitably linked with the "Diabeta" database in respect to the evaluation of overall outcome (e.g. cost) of health care related to diabetes, or its complications, for our patients.
2. Visits to the NHS organisations outside St Thomas' Hospital, such as local Health Authorities, General Practices, Communities Care settings and Social Services councils. This was to investigate the feasibility of linking secondary/primary care information systems with "Diabeta" for evaluating the outcome of care in diabetic patients.
3. Investigation of the NHS-Wide Clearing Services
   Meeting with Paul Eveson, Policy Implementation Manager, Headquarters, the Department of Health in Leeds, to understand the way contract data set transfers from health care provider to purchaser, and also to find out if there is a national hospital episode database available in UK.
4. Investigation of records system in Prescription Pricing Authority (PPA)
   Meeting Mark Spencelayh, PPA system development manager in Newcastle Upon Tyne, and discussing the possibility of linking the prescription resource to "Diabeta".
5. Selecting a cohort of patient records as a cohort for the pilot study.
6. Sending the cohort records to all relevant information systems within the NHS.
7. Recommendation on the feasibility of linking "Diabeta" with other relevant information systems within the NHS.

Chapter 5

Patients and Methods

5.1 Flagging "Diabeta" records in the NHSCR record system

When this present study was initiated, over 7,000 diabetic patient records were available in "Diabeta" since 1973. Patient demographic data (forename, surname, date of birth, gender, hospital number, last address and date last seen in diabetic clinic) required by the NHSCR, were retrieved from "Diabeta" (the Office of Population Censuses & Surveys). This information was then formatted onto floppy disks according to the NHSCR regulations (subject to special processing for data security) then mailed to the NHSCR to be updated. Records were automatically matched with the NHSCR database using the first four characters of each patient's surname and forename and with their date of birth. If all these items were consistent with analogous items on record in the NHSCR system, a match and link was considered to have been made. If there was more than one matching data set in the NHSCR records, automatic matching alone was not successful. Such records would be skipped and the operator would carry on with the searching process by using other identification data (such as gender, address and last date seen in the diabetic clinic). Patients' addresses were not recorded in the NHSCR, but the NHSCR operator could use the patients' addresses for comparison with the FHSA area with which patients were registered. A "trace model" was used to estimate the probability of both records belonging to the same person. In the case of poor quality identification data or patients who died or emigrated before 1991, record matching was done manually using the old record systems.

When patient records were matched, and linked, with records in the NHSCR, a unique study number was assigned. This study number (only disclosed to the authorised medical researchers in the NHSCR) had been entered (or 'flagged') into the NHSCR records. Therefore, whenever a patient death occurred, a copy of the death certificate would be forwarded to us within a few weeks. Details of 'exits' from the NHS (such as embarkation) would also be notified. Meanwhile, the corrected discrepancies (name and date of birth) and patient identifiers (FHSA code and NHS number), not recorded in "Diabeta", were also provided (the Office of Population Censuses & Surveys) (Figure 5.1).

The handling of patient data on "Diabeta" and the NHSCR complied with the Data Protection Act (1984). In the case of record linkage between the NHSCR database and "Diabeta", the NHSCR have issued specific security measures for ensuring transfer of authorised data (Thompson EJ et al 1993).

Results of updated records are currently provided in paper form. The NHSCR plans to provide these results electronically (on floppy disk) in the future. Data on floppy disks would then be subject to data security processing (Thompson EJ et al 1993).

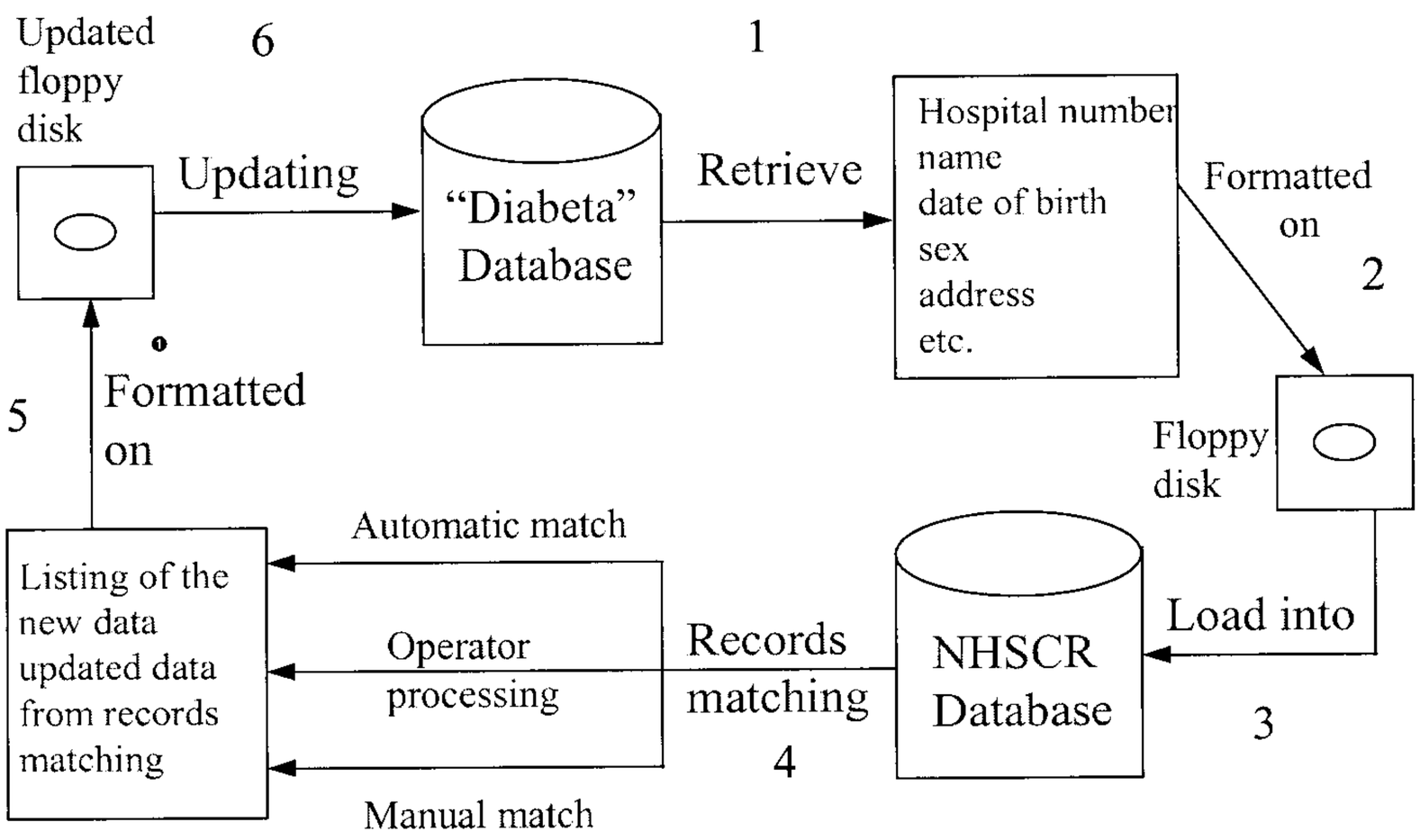


Figure 5.1 Protocol of the record linkage system between "Diabeta" and the NHSCR database.

The transference of updated patients demographic data from the NHSCR to "Diabeta" can be performed on a regular monthly basis. New patient records can be retrieved from "Diabeta" and sent to the NHSCR for prospective "flagging" on a similar regular basis.

5.2 Tracing patients records from a cohort in other relevant information systems within the NHS

In order to examine the feasibility of linking "Diabeta" with other health service information systems within the NHS, a cohort of IDDM patients (age onset of diabetes <40 & with treatment of insulin) were selected as a representative cohort for the pilot study. All patients from the cohort first attended in the Diabetes Clinic at St Thomas' Hospital between 1982 to 1985, and were re-examined in 1995. This study is mainly concerned with IDDM, since suffering from this type of diabetes are more likely to present with complications such as neuropathy and nephropathy in their early ages (Leese B et al 1992). This study also intends to evaluate the possibility of using "Diabeta" to build a cost-model so as to assess the cost of treatment of IDDM in UK, and compare them with an established Global Pharmaco-Economic Research (e.g. The economic study in IGF-1 treatment in IDDM in Gentech, US).

These records, formatted onto floppy disk/paper with patients' identifiers (hospital number, NHS number, name, date of birth, address, postcode), were sent to the manager of each relevant information systems with a request to carry out a record matching procedure (automatic or manual).

The number of matches and amount of information available was recorded and analysed. This was to estimate the proportion of records that can be matched on each relevant system, and which type of resources were more consumed by the diabetic patients. This helps to determine which resources are more sensibly included into a cost-modelling study.

Chapter 6
Statistical Analysis

Record linkage between "Diabeta" and the NHSCR has provided an updated patient vital status (alive/deceased) for the entire "Diabeta" database. Information on mortality in our diabetic population is now available, based on this updated database. In this section, the statistical technique used in the mortality analysis will be shown.

6.1 Standardised Mortality Ratios

The overall comparison of mortality in the cohort of diabetic patients in "Diabeta" applied to the general population of England and Wales, was carried out by calculating the ratio of the observed number of deaths, to expected number of deaths, i.e., the Standardised Mortality Ratios (SMRs). The expected number of deaths was calculated using the so-called "person-years" method. The SMR is the ratio of the number of deaths observed (D).

$$D=\sum_{j=1}^{J} dj$$

in which d_j denoted is the number of deaths observed, E* is the expected number of deaths multiplied by 100 (assuming that diabetic patients have the same mortality rate as the standard population, and "j" is the age in years). In the calculation of expected death, each person is assumed at risk up to the date of the analysis, the date of death, or the date the person was lost to the follow-up procedure (whichever is first).

$$\text{SMR} = \frac{D}{E^*} \times 100 = \frac{\sum_{j=1}^{J} d_j}{\sum_{j=1}^{J} n_j \lambda_j^*} \times 100, \text{j} = 1,2,... \qquad (6.1.1)$$

A SMR greater than 100 indicate excess mortality in a given cohort as compared with the standard population. The expected number of deaths in our patients was estimated by multiplying the number of person years at risk (n_j) by the reference death rates (λ^*_j) (Berry, 1983).

Confidence intervals for SMRs were calculated using the *Poisson* assumption. The confidence limits were obtained by first finding lower (*L*) and upper (U*)* limits μ_L and μ_U respectively for the mean [$\mu = E(D)$] of the *Poisson* distributed observation D (the number of deaths observed).

Where: $$\mu_L = D\left(1 - \frac{1}{9D} - \frac{Z_{a/2}}{3D^{1/2}}\right)^3$$

And: $$\mu_U = (D+1)\left(1 - \frac{1}{9(D+1)} - \frac{Z_{a/2}}{3(D+1)^{1/2}}\right)^3 \qquad (6.1.2)$$

The confidence interval for SMR are then calculated $SMR_L = \mu_L / E^* \times 100$ and $SMR_\mu = \mu_U / E^* \times 100$.

$Z_{a/2}$= 1.96 for 95% confidence intervals. If the confidence limits covered the value 100, the null hypothesis is accepted, hence the cohorts studied did not experience significant excess mortality compared with their reference populations.

6.2 Modelling of mortality data

Initial covariates measured at first visit naturally change over time. Therefore the Cox proportional hazards model, which is time dependent, does not fit our data and the Accelerated Failure Time model (ACF) (Collett D 1994; Kalbfleishch JD et al 1980; Wei LJ 1879) was used to assess the effect of the prognostic variables on mortality. The ACF was chosen as any variable that would be expected to 'accelerate' a patient's risk of death. A description about the model follows. Let T_i (I=1,2, ..., n) represent the failure time for the i'th patient and x_i a prognostic variable for that patient. It should be noted that T_i consists of failure times and censored observations. When a study has ended, many patients are still alive and so, not all observations are failure times. These readings are important as they provide information on survival and are thus censored, indicating that they are a measure of survival. The ACF model assumes a linear relationship between the log of T_i and x_i, that is

$$\log(T_i) = \beta x_i + \varepsilon_I,$$

Where ε_I independently and identically distributed follows a distribution F, known as the survival distribution functions. The value of β corresponds to the effect of the prognostic variable. The above model is similar to a simple linear regression (s.l.r.) model except that the errors, ε_I, do not follow a normal distribution with mean 0 and variance.

Two commonly used measures in survival analysis, are the survival function, and hazard function. The survival function, s(t), is the probability that the survival time of a patient is greater than or equal to t. The hazard function, h(t), is the probability that an individual dies at time t. For the data in question a number of survival distributions were considered. However, by analysing residuals and other model diagnostics. The Weibull survival distribution was found to be the better choice (by analysing residuals and other model diagnostics). Based on this model, patient I, with a prognostic variable reading x_i will have survival and hazard functions at time t, of

$$S_i(t) = \exp[-a\exp(\beta x_i)^{bt}] \qquad h_i(t) = [\exp(\beta x_i)]^b abt^{b-1} \qquad (6.2.1)$$

where a and b are 'shape parameters' for the Weibull survival distribution. In the above, the quantity $\exp(\beta x_i)$ is called the acceleration factor, and is denoted by ϕ. When $\phi>1$, this indicates an acceleration, and the hazard function increases, whilst the survival function decreases. The opposite can be stated if $\phi<1$. For the data presented, the time to failure is taken as the age of death of a patient. In order to find which prognostic variables significantly influenced mortality, a stepwise procedure was used to accept or reject variables (at the 5% significance level) from the ACF model.

A commonly used measure to look at differences between values of a prognostic variable is the proportional odd ratio. This can be quoted if the survival functions for two groups of patients are parallel, such that the odds ratio is $\varphi = \phi^b$. However, for the data and survival model in question, the survival times are not proportional and the odds ratio should not be used as they may be misleading. Instead, the acceleration factors are investigated. Let us consider two patients (A and B) with acceleration factors $\phi_1 = \exp(\beta x_1)$ and $\phi_2 = \exp(\beta x_2)$ respectively. The different factors can be used

to compare the acceleration for both patients. This is done as a ratio that is estimated by

$$\exp[\beta(x_1-x_2)] \qquad (6.2.2)$$

If this ratio is greater than unity, then patient A has a higher acceleration factor than patient B and therefore has a smaller survival time. If the ratio is less than unity the opposite can be stated. The approximate 95% confidence interval for the estimated ratio is obtained using

$$\exp\{(x_1-x_2)\,[\beta \pm 1.96 \times \text{s.e.}(\beta)]\} \qquad (6.2.3)$$

The above is based on one prognostic variable; if there are more variables, then all but the variable of interest are set equally for both patients. The ratio obtained in this manner represents the acceleration factor for patient A relative to patient B, adjusted for the other prognostic variables.

Chapter 7
Geographic Information System (GIS) used in the analysis

In order to use the data obtained from the record linkage ("Diabeta"-NHSCR) more efficiently in terms of our epidemiological studies, a link with a GIS software was considered. This idea was to analyse socio-economic factors for excess morbidity and premature mortality. The techniques of linking patient's clinical data in "Diabeta" with a geographic data set will be discussed in this chapter.

7.1 Linking patients' clinical data with geographical data

The author used MapInfo software (http://www.mapinfo.com) to link and access the "Diabeta" database. MapInfo can use one postal address together with postcode-correlated data from many different databases (the "Diabeta" database, census database etc.) to bring the various elements of the data into a single-view geographically as a map, and to illustrate important relationships, that would otherwise remain invisible.

Initially the author used a basic Greater London map that included Electoral Ward boundaries together with 1995 National Census data for each ward. This base-map was then overlaid with other data layers such as Roads, Bus routes, Underground routes, General Practices, Chiropodists, Pharmacies and Health Authority boundaries.

The author then used 'geo- patients' postcodes/postcode sectors, which assign an 'x' & 'y' reference to plot on the map as a layer overlapping the patients' clinical database with an Electorate Ward map (as described above). The software will automatically link patients' postcodes with the relevant ward together with its census data (Figure 7.1).

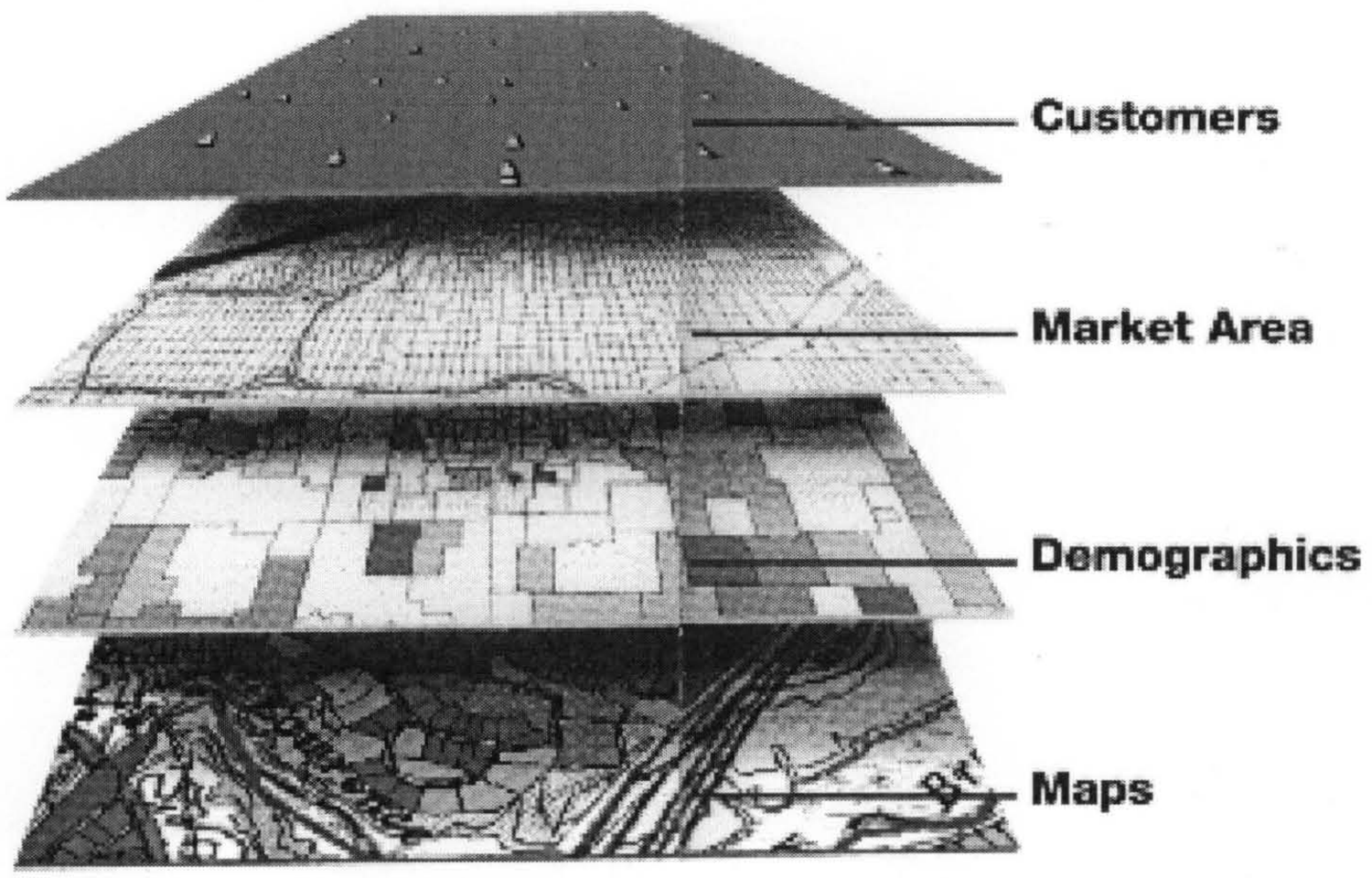


Figure 7.1 A various layers related to different information overlapping in a map.

The maps will also be able to show the distribution of diabetic patients (by any recorded variable) within the LSLHA area (Figure 7.2).

7.2 Analysis of data using GIS and standard statistical technology

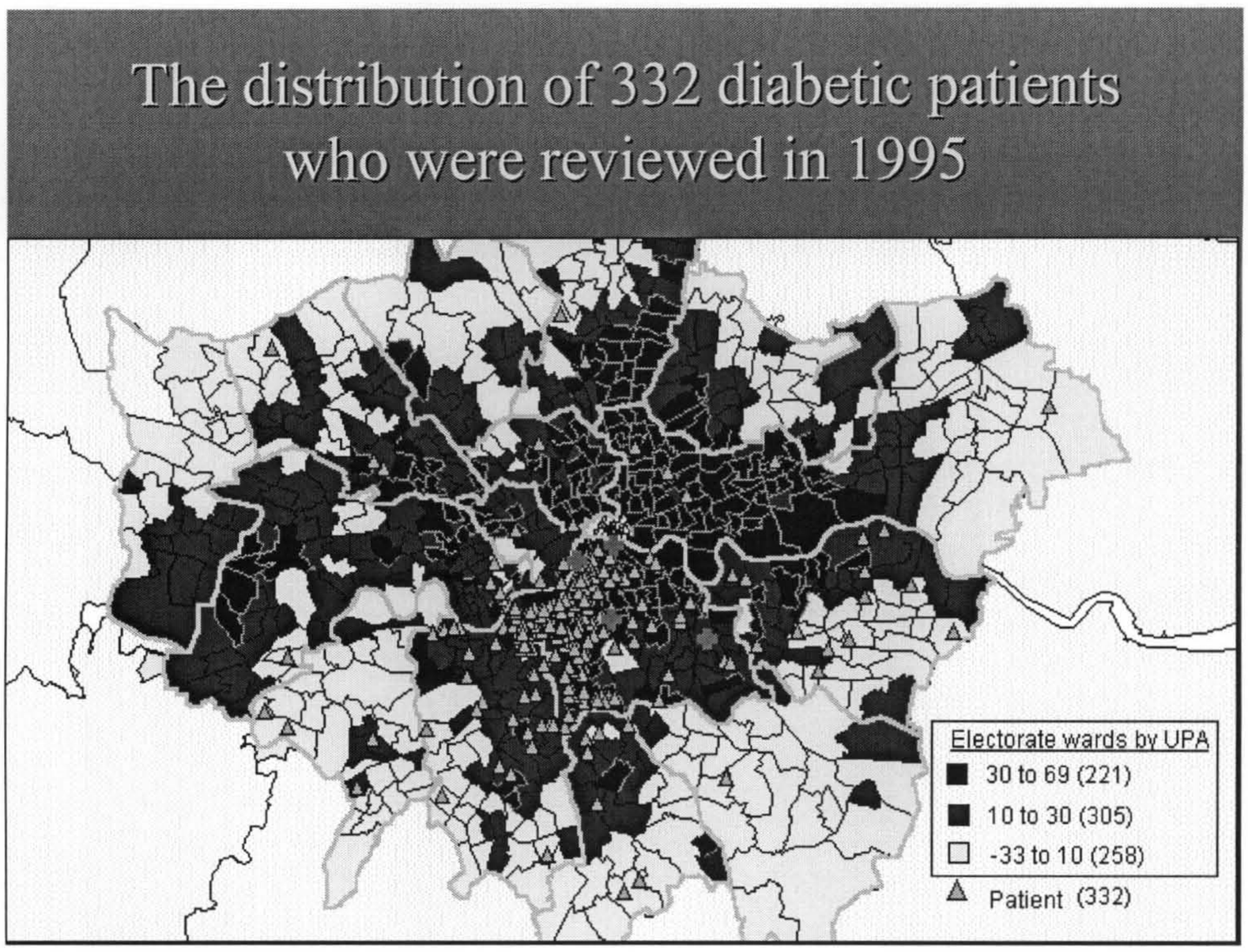
The purpose of this analysis is to identify socio-economic and environmental factors related to diabetes outcomes.

There are several established indices, which are based on the census data listed below.

- SCOTDEP - developed by Carsairs and Morris for analysis of Scottish health data.
- TOWN - as used by Townsend at al in an analysis for the Northern Region.
- JAR- Jarman and associates developed this measure in relation to need for primary care services.
- DOE - Department of Environment measure, developed mainly in relation to urban policies.
- SDD - Scottish Development measure, developed mainly in relation to urban policies.

The most popular indices of deprivation, are Jarman and Townsend indices.

Using MapInfo, the distribution of social deprivation within Greater London can be shown on a map overlapping patients' addresses. Standard statistical methods (χ-test and t-test) were used to define the correlation between social deprivation and diabetes outcome measures.



■ 30 to 69 = lower socio-economic areas

-33 to 10 = high socio-economic areas

Figure 7.2 A map showing distribution of a cohort of diabetic patients in Greater London with a background of health authority boundaries and Jarman Under Privileged Area categories.

Section Three

Results

The benefits of the record linkage between "Diabeta" and the NHSCR were estimated in four studies.

The first was a 'flagging study' in the NHSCR seeking to estimate the possibility of a linkage between the NHSCR and "Diabeta", together with the costs incurred.

The second was to analyse how such a linkage can be used for medical research and epidemiology studies. The excess morbidity, premature mortality, and risk factors related to mortality of diabetes were assessed, based on the updated data from "Diabeta". In addition the predicted value on outcome of diabetes was measured also via the updated "Diabeta" database.

In the third study, benefits of the records linkage between "Diabeta" and the NHSCR were explored by linking the existing record linkage to a Geographic Information System (GIS), in an attempt to identify which social and geographic factors may be related to outcome in diabetes care.

The efficiency of the linkage is illustrated in the final study. The question of how often, and how accurately patients' vital status can be updated on "Diabeta", was addressed.

The feasibility of implementing record linkages between 'Diabeta' and other relevant information systems within the NHS was based on a sample study. This was a 'tracing study' on all relevant information systems intending to estimate the benefits of a further electronic linkage between two systems, and also the costs incurred.

Chapter 8

Assessing the possibility of the linkage between "Diabeta" and the NHSCR

All patients registered on "Diabeta" at the time of study (April 1995) that had local identifiers were formatted onto the floppy disk and sent to the NHSCR.

NHSCR researchers carried out 'tracing' and 'flagging' studies on the patient records supplied. The records provided by "Diabeta" were matched with records in the NHSCR, and discrepancies corrected whenever detected. As a result, "Diabeta" records were updated and the results provided in paper form. Patient's identifiers not recorded in "Diabeta" records, but available in the NHSCR records, were also provided in paper form. An "MS Access" database was designed and built to enter all updated patients' local demographic data received from the NHSCR. The hospital number was used to retrieve records in the database. This procedure will continue until the results from the 'flagging study' in the NHSCR become readily available on a magnetic format (floppy disk).

The patients' identifiers held in the NHSCR are illustrated in Figure 8.1. The shadowed identifiers are those which are also recorded in "Diabeta".

NHS No.	Date of last registered with FHSA FHSA code	Surname	Forename	Date of Birth	Address	Date of seen in DEDC	Hospital number
Supplied by ONS		*Identified by ONS*					

supplied by "Diabeta" Database at St Thomas' Hospital

Figure 8.1 Patients' identifiers held in the NHSCR and "Diabeta"

During this study the following information was supplied to the author in paper form and entered onto the "MS Access" database:

1. The trace possibility
2. The updated patient's vital status for whose records could be traced in the NHSCR and copies of death certificates for the deceased patients.
3. Old NHS numbers were provided for all patients' records in the cohort that could be traced in the NHSCR. After a replacement of the new NHS number on the NHSCR records in 1995, a new NHS number has been received for those whose registration status changed since their records were 'flagged' on the NHSCR.
4. For patients alive who could be traced by the NHSCR, the code of their current FHSA and the date of registration with that FHSA were also provided.
5. The identifiers of all patients that could be traced were re-identified in the NHSCR and corrected. Any discrepancies with corresponding details in the NHSCR database were notified.
6. The NHSCR traced patient records using several available procedures according to the quality of the identifying data provided. The number of records identified with each specific procedure was also reported by the NHSCR.
7. Based on the above results the NHSCR estimated the total cost of 'tracing' or 'flagging' on the NHSCR database for all patients registered in the "Diabeta" database.

The results of the study with the NHSCR are explained and discussed under the following headings:

8.1 The possibility of a linkage between "Diabeta" & the NHSCR

7,542 diabetic records in "Diabeta" (up to 1994) were submitted to ONS for tracing on the NHSCR. Of these, 6,851 records (91%) in "Diabeta" were successfully linked with records in the NHSCR (i.e. there was enough evidence to indicate valid matched ('linked') records belonging to the same patient). 655 records (8.7%) failed to match any existing records in the NHSCR. 36 records (0.48%) could only possibly be linked if complete patient identification information was supplied to the NHSCR (Figure 8.2).

Total patient records	Traced (linked) records	Possible traced (linked) records	Untraced (Unlinked) records
7542	6851 (91%)	36 (0.48%)	655 (8.7%)

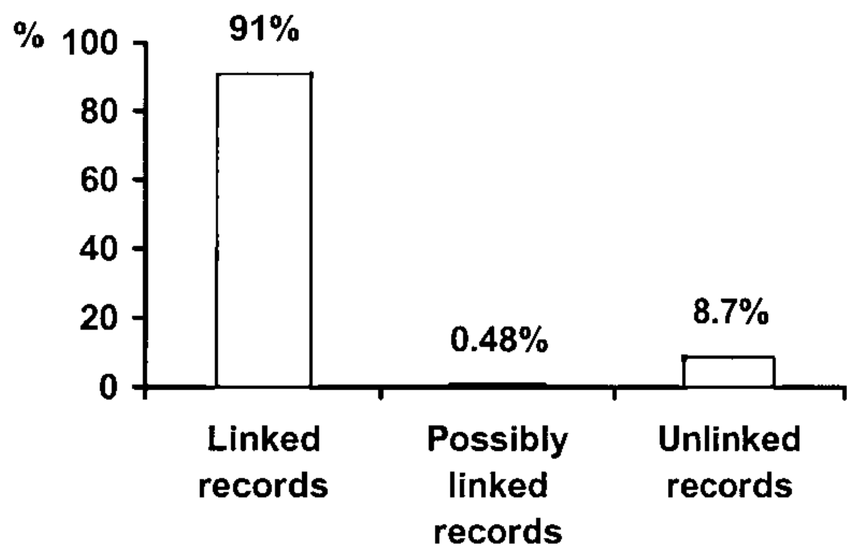


Figure 8.2 Linkage of records in "Diabeta" with the NHSCR database

In conclusion, the majority of records in "Diabeta" were successfully traced and flagged on the NHSCR database.

8.2 Updating the patient's vital status (alive/deceased) in "Diabeta"

The NHSCR provided information on the vital status of 91% patients submitted, which resulted in a successful trace of 1, 670 deceased patient's records. Only 239 (14%) of these patients was notified to "Diabeta" (Figure 8.3).

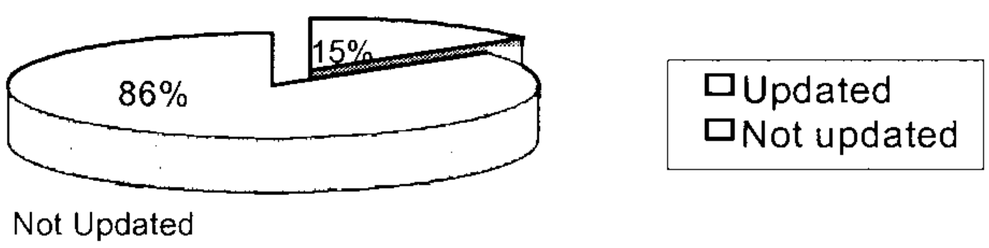


Figure 8.3 Updated death records in "Diabeta".

86% of diabetic patient deaths had not been notified to the hospital and hence was not recorded in "Diabeta" prior to the linkage. Mortality analysis in a cohort of our diabetic patients would therefore be meaningless without a "Diabeta" - NHSCR linkage.

8.3 Provision of FHSA code for all patients alive traced in the NHSCR

Out of those linked (or matched) records (n= 6, 851), nearly 71% of our patients were still registered with a NHS doctor (Figure 8.4). For all these patients, codes of their current FHSA and the dates of registration were provided by the NHSCR, enabling us to trace patients' current GPs through the relevant FHSAs.

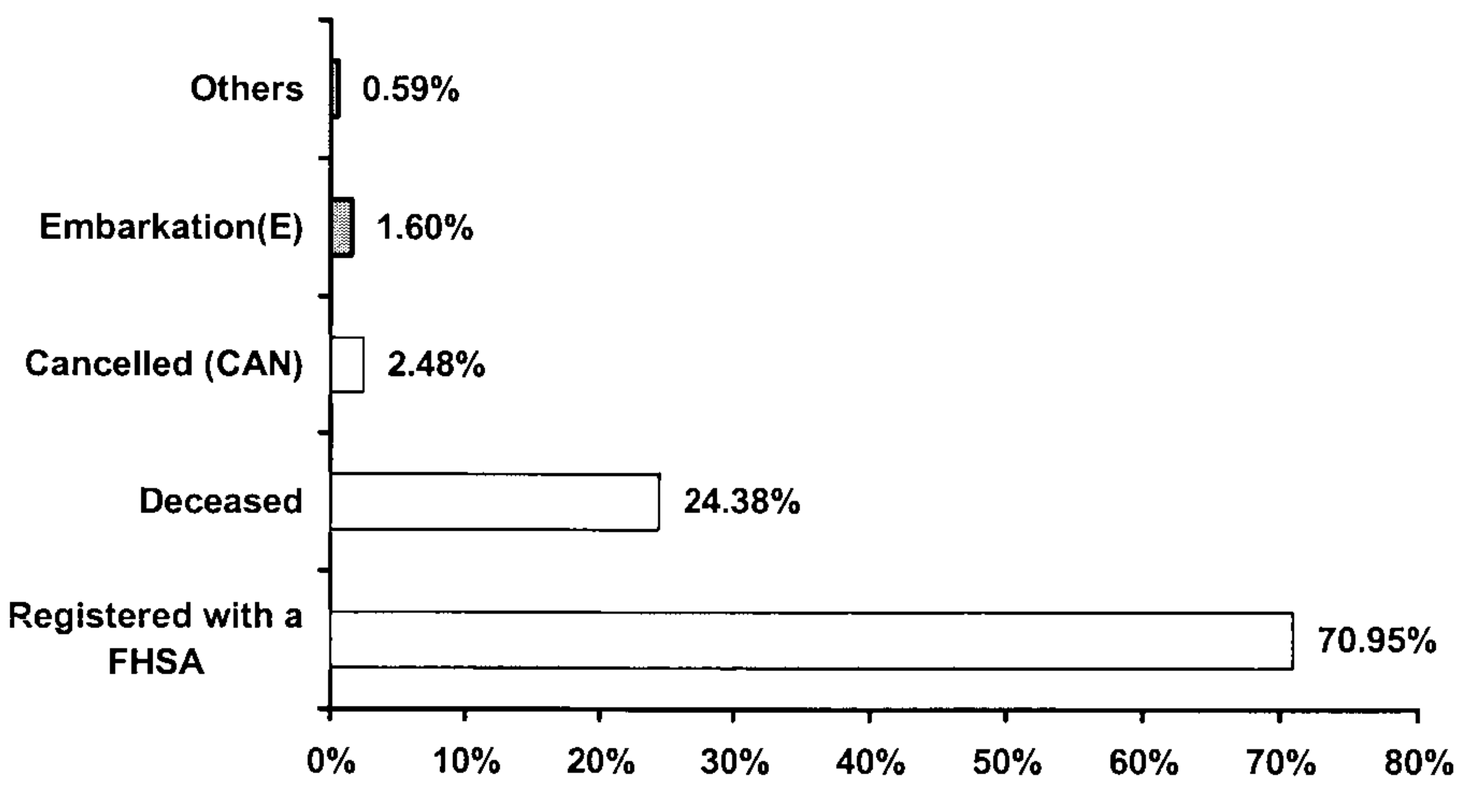


Figure 8.4 Registration status of diabetic patients in the NHSCR database.

320 (4.7%) patients were no longer registered with the NHS and their records would be marked by a different code (e.g. embarkation etc.). In the event of a death, within the UK of patients no longer registered with an NHS doctor (e.g. long-stay psychiatric patients, prisoners and people served in Arm Forces), details of death will still be notified to "Diabeta" when it occurs.

8.4 NHS number

"Diabeta" has a field for the NHS number, but only recorded <1% of cases as the number has, in the past, been rarely used by GPs in their referral letters. The NHSCR has provided NHS numbers on all our patients and this has now been added to the "Diabeta" database.

8.5 Updating patient identifiers in the study

As already discussed, the 'tracing' and 'flagging' study performed by the NHSCR on patient records extracted from "Diabeta", was based on matching these records with corresponding records in the NHSCR based on patients' identification data (e.g. surname, forename, initial and date of birth etc.).

In some cases discrepancies between these identifiers were detected, recorded in paper form, and sent to the author - the results of which are displayed in Table 8.1. It shows the amount of patient's identification data disagreed between two sets of records ("Diabeta" & NHSCR). The proportion of the disagreement was compared between successful matched and possible matched records.

Table 8.1 Results of analysis on disagreement of cohort patient's identifiers in pairs of **linked (or matched) records** and pairs of **possible linked (or matched) records** between "Diabeta" and the NHSCR.

Patient identifiers that were in disagreement in the two sets of records (NHSCR & "Diabeta")	Number of pairs of linked records* Total (6851) % of pairs linked records	Number of pairs of possible linked records* Total (36) % of pairs of possible linked records
Surname	144 2.1%	5 13.8%
Forename	500 73.3%	10 27.8%
Initial name	1355 19.8%	15 41.7%
Date of Birth	69 1.0%	26 72.2%

The author concludes that discrepancies between patient's identifiers in "Diabeta" and the NHSCR database appeared proportionally higher in the 'possible linked (or matched) records'

The highest percentage of disagreement found in 'possible linked records' appeared in the date of birth. This could be interpreted as follows: the more accurate the Date of Birth provided by "Diabeta", the higher the probability of successful record linkage.

8.6 Numerical distribution of different records matching procedures

Depending on the quality of patient data supplied to it, the NHSCR uses different procedures for matching records. In this case, quality of the data from "Diabeta" was sufficiently good to enable 42% of the records to be matched on the NHSCR database; 39% were matched on the NHSCR database 'semi-automatically' (requiring operator interference) and 19% were matched manually on the old record systems (Figure 8.5).

Matching procedure	Cost per record (£)	Number of traced records (%)	Total cost (£)
Automatic matching using computerised system	0.30	2 877 (42%)	863.1
Operator interference	1.70	2 672 (39%)	4,542.4
Manual matching	4.75	1 302 (19%)	6,184.5
	Total	6 851	11,590

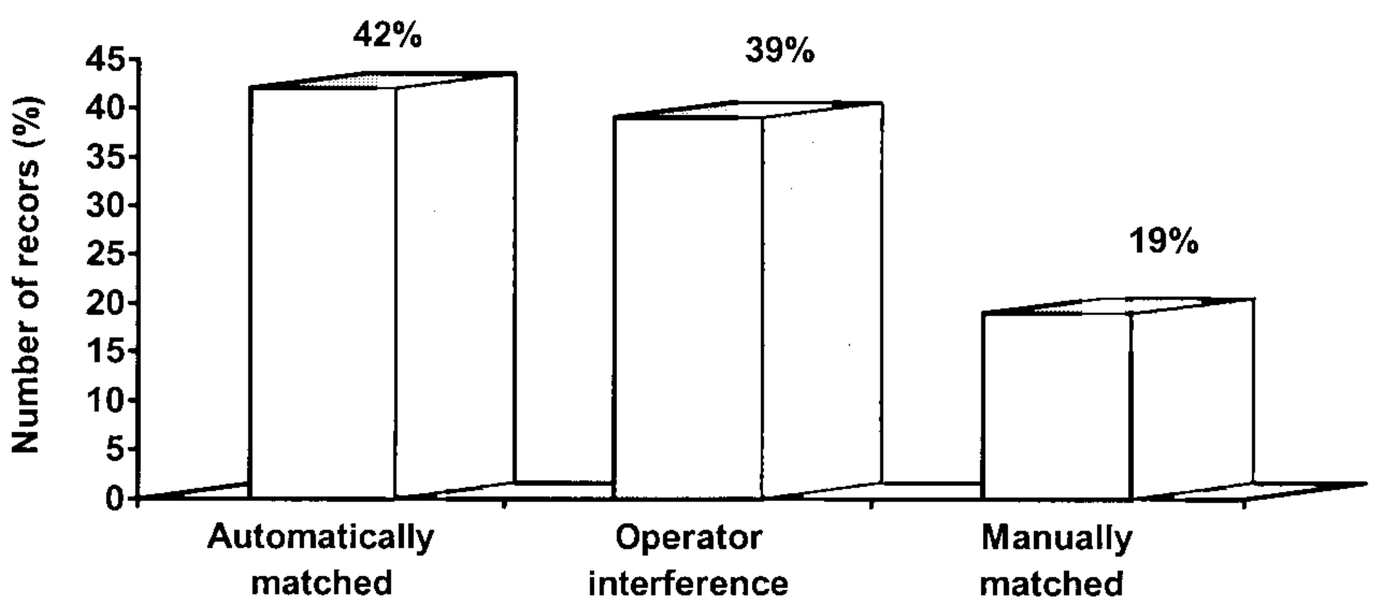


Figure 8.5 Procedures used in record matching

The cost of automatic matching of records was significantly lower than that of manual matching. If the NHS number had been recorded on "Diabeta", the cost-savings for the electronic linkage between "Diabeta" and the NHSCR database would be significant. The average cost of 'flagging' events (e.g. death certificate, embarkation or changed FHSA registration etc.) on the NHSCR database and retrospectively or prospectively notifying changes on patients' demographic or

NHS registration status was between £1 and £2 per patient. There are nearly 400 new patients attending the diabetic clinic each year. Therefore the recent annual fees on the 'flagging' study will be about between £400 and £800. This is a 'one-off' charge and is not seen as expensive considering the importance of having an accurate and complete database that would enable us to stop sending follow-up appointments to people who have died, and that can support meaningful epidemiological studies and reliable clinical audits. These costs can be even further reduced if patient records on "Diabeta" can be automatically matched with those of the NHSCR. Better quality and quantity of patient identification information on "Diabeta" would lead to more advanced and effective matching techniques with less cost.

8.7 Estimation of the efficiency of updated data received in "Diabeta" through the NHSCR

As previously mentioned, the notification of the NHS registration status (death, embarkation and change of health authority etc.) in the NHSCR, is collected by "Diabeta" through a 'flagging' study. The notification of changes of patients' registration status is downloaded to the NHSCR database weekly via an electronic link from each FHSA (a Racal Link) or via a magnetic media (e.g. floppy disk) from other organisations (e.g. Local Register or Home Office) weekly. The maximum time acceptable for notification after an event is still however uncertain.

In order to examine how long the notifications can be received as compared to the time of events occurred, the author randomly selected a cohort (n=100) from those patients who had notified their change in the NHS registration status, after their records were 'flagged' on the NHSCR. The average number days between date of events occurrence and receiving by author was calculated to an average of 75 days (2.5 months) (Table 8.2).

Table 8.2 Estimation of the efficiency of receiving notifications from NHSCR to "Diabeta".

Event (n=100)	Originally Recorded in	Notifying to NHSCR		Notifying to "Diabeta"
		Media	Length	Length (days)
Death (n=55)	Local Register	floppy disk	3 weeks	77
Changing current FHSA (n=30)	FHSA Patients Registration System ('Exeter')	network	1 week	92
No longer registered with a NHS doctor (n=10)	Long stay psychiatric institutions, army force, and prisons	paper form	unknown	68
Embarkation (n=5)	Home Office	paper form	unknown	86

The NHSCR generates reports of the 'flagging study' for each researcher every three months. Although it can receive notification of updated FHSA registration status within a week, notification to "Diabeta" can take up to three months. This delay can be significantly reduced if "Diabeta" is connected to the NHSCR via a network. Further discussion on this issue is in the last section.

8.8 Conclusions

Most records in "Diabeta" could be linked with the NHSCR and such a link can 'automatically' notify deaths to "Diabeta" in a cost-effective way. The benefits of this update are considerable: (a) 86% of diabetic patient's death had not been notified to the hospital and not recorded on "Diabeta". Mortality can now be assessed accurately as an outcome measure in our diabetic population. (b) Provision of the NHS number to "Diabeta" (not available previously on most patients seen in the hospital) is a key for exchanging information within the NHS-wide network. (c) The geographical location of patients on the database was updated, enabling tracing of patients for long-term studies and analysis of movement. (d) Unnecessary distress of patients' relatives by sending appointment letters to deceased patients can be avoided. The benefits of such a link on epidemiology studies are discussed in the following Chapter.

Chapter 9

Assessing excess mortality in diabetic patients at St Thomas' Hospital (1980-1994)

Diabetes is associated with mortality rates 2 or 3 times higher than the rest of the population (McCarty D et al 1994). Various studies have shown that, although patients usually die of the complications of diabetes, the term 'Diabetes Mellitus' does not always appear on the death certification. Statistics based on death certificates alone greatly underestimate the true incidence of diabetes-related mortality. The record linkage between '"Diabeta"' and the NHSCR has enabled the diabetic clinic at St Thomas' Hospital to ascertain all deaths in the diabetic population attending the hospital and to access their causes of death from death certificates. A mortality study based on such a population will be more accurate and meaningful than a study based on death certificates alone. In this Chapter, the true incidence of mortality in our diabetic population is assessed, and the overall mortality in diabetic subjects is compared with that of the general population.

For our patients who are 'flagged' in the NHSCR database, the NHSCR requests death certificates (Appendix B) from Titchfield Health Authority for those deceased and these are sent to the author. The following information was supplied on paper form.

1. Local birth and death registration data to indicate in which local register patients registered.
2. Patients demographic data included: patients' name, gender, maiden name, date and place of birth, date and place of death, occupation and address.
3. The information of the informant is recorded in the name and address of the person who is responsible for death registration. The relationship between informant and deceased patient is also described.
4. The causes of death are recorded as six entries in Arabic numerical order as an *original multiple cause of death*. The first three entries describe *the main cause of death* (Part I). They are weighted and compiled from first to third. The second three entries are the *contributed cause of death* (part II). The *underlying cause of*

death is extracted from multiple cause of death. The causes of death are coded by the *International Classification of Disease, Revision 9 (ICD-9).* The underlying cause of death has been used for the ONS annual statistical report on mortality in districts and regions in England and Wales. The WHO has also used it for mortality studies since 1979.

5. The NHS number (both old and new, have been recorded).
6. The date of registration and signature of informant and the doctor who certified the certificate.

All this information has been entered onto an MS Access database. The most valuable information obtained from the linkage between "Diabeta" and the NHSCR via death certificates are the date of death, cause of death and place of death; such information was previously not recorded in "Diabeta". The following studies have shown how this information can be used for assessing excess mortality in diabetic patients, whilst trying to answer these questions:

1. Why is 'Diabetes Mellitus' underestimated on death certificates?
2. Do such deficiencies affect the estimation in overall mortality?
3. What are the trends in all-cause diabetes mortality?
4. Is diabetes an independent predictor of mortality (overall and for selected causes of death) in both men and women?
5. Do diabetic women retain a relative advantage in mortality?
6. Which risk factors (clinical/social) are related to excess mortality in diabetes?

The aim of this study is to compare the findings based on the updated (linked) "Diabeta" database, with other previous mortality studies. As part of this thesis, the author intends to evaluate how valid epidemiology studies can be achieved on the "Diabeta" database updated through the record linkage techniques.

9.1 Mention of 'Diabetes Mellitus' on death certificates

Diabetes mellitus has been called 'an underestimated public health problem'. Partly because it is known that diabetes (being a chronic disease) is associated with other conditions such as ischaemic heart disease, which is more 'visible' to the certifying physician and thus selected as the underlying cause of death, often omitting the term 'Diabetes Mellitus' (Fuller JH et al 1983). Various studies report considerable variation among physicians in reporting the diagnosis of diabetes on death certificates of people with diabetes (Kleinman JC et al 1988).

In our study, out of 1,670 obtained death certificates, 'Diabetes Mellitus' (ICD 9 codes 250.0-250.9) was mentioned as an underlying or contributory cause of death in only 36% of death certificates (Figure 9.1) - 40% before 1980, 40% between 1981 and 1985, 38% between 1985 and 1990, 35% between 1991 and 1995. These figures are similar to those reported by Fuller JH (Fuller JH et al 1983).

For deceased diabetic patients, the underlying cause of death was more likely to have 'Diabetes Mellitus' mentioned in Part II (*contributed cause of death*) of the death certificate (23.74%) than Part I (*main cause of death*) (12.23%). Fisher's exact test, two-tail $p<0.05$).

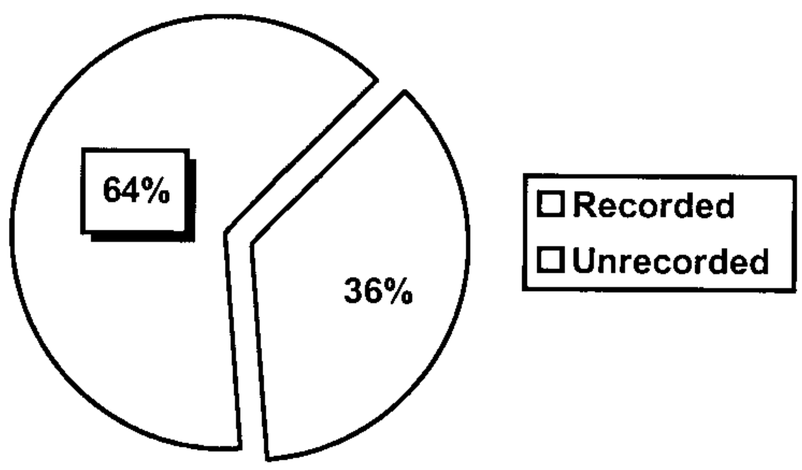


Figure 9.1 Mention of 'Diabetes Mellitus' on death certificates

Women were more likely to have DM mentioned than men (39% vs. 37%), although this difference is not significant. The rate of mention of DM is also related to the cause of death (Table 9.1). For females, a relatively higher proportion of death certificates has DM mentioned when the underlying cause of death is

ischaemic heart disease (ICD 9 code: 410-414). Males were more likely to have DM mentioned when the underlying cause of death is cerebrovascular disease (ICD 9 code: 430-438). The present study proves that patients who did not have 'diabetes' listed on the death certificate were more likely to die from a malignancy or accident. This has also been demonstrated in other studies (Ochi JW et al 1985). DM is more likely to be mentioned on death certificates in subjects <44 years (69% 15-44 yrs, 38% 45-64 yrs, 36% 65-74 yrs, 35% 75-84 yrs, 40% 85+ yrs), although these differences were not statistically significant. In the Fuller JH study (1983), the proportion of certificates mentioning diabetes as the underlying cause of death declined with age from 70% in subject <15 years to 20% in subject > 65 years.

Table 9.1 Proportion of patients who had DM mentioned on the underlying causes of death in both male and female groups.

Underlying causes of death (ICD 9 code)	Men		Women	
	Number of certificates	Mention of DM n(%)	Number of certificates	Mention of DM n(%)
Cerebrovascular disease (430-438)	82	32 (39)	76	25(33)
Diabetes (250)	106	106(100)	111	111(100)
External causes (E800-E999)	10	1(10)	6	1(17)
Ischaemic heart disease (410-414)	269	75(28)	168	61(36)
Neoplasm (140-239)	162	35(22)	143	20(14)
All other cause of death	227	65(29)	184	51(27)
All causes of death	856	314(37)	688	269(39)

Renal failure (chronic, acute or end stage (ICD 9 code: 584, 585 and 596)) was mentioned as an underlying and contributory cause of death on 156 death certificates (10%). Certificates that mentioned renal failure (on any entry) were more likely to mention diabetes (68.15% vs. 41.48%, Fisher's exact test, two-tails $p<0.003$). This phenomenon has been previously recognised in a Danish study (Green A et al 1983). Hypoglycaemia (ICD 9 code: 2510) was mentioned on only 14 death certificates (0.9%) (nine men, five women).

In conclusion only one third of death certificates mentioned DM, because the primary cause of death gives precedence (such as ischaemic heart disease,

cerebrovascular disease or renal failure) and "diabetes" gets forgotten or overlooked. DM is more commonly mentioned on death certificates in women than men. Valid studies on diabetes related mortality in UK were not possible using death certificates alone. Underestimation of diabetes-related deaths in the diabetic population will be further discussed in the following sub-section.

9.2 Overall mortality in a cohort of diabetic patients at St Thomas' Hospital

This study compares diabetes mortality rates on a cohort of patients extracted from the updated "Diabeta" database, with the mortality rates determined on death certificates alone. The cohort of patients used for this study are those whose records were 'flagged' in the NHSCR by 1994.

Mortality rates can, of course, vary when based on the definition of a diabetes-related death. Consequently, the author calculated the rates in three different ways from death certificates.

(1) Counting deaths on those who have 'DM' mentioned on underlying cause of death.
(2) Counting deaths on those who have 'DM' mentioned on contributing /underlying cause of death.
(3) Counting all deaths that known as diabetes, regardless of the death certificate information.

Diabetes mortality rates using these different definitions are displayed in Table 9.2. It is calculated in the usual manner by dividing the number of deaths by the sum of individual follow-up years in each sex and age group (person-year method). For the underlying cause of death, mortality rate was 3.4 per 100 person-years based on 7, 542 diabetics at risk. If a diabetic death were taken to mean any death where diabetes was mentioned on the death certificate, the mortality rate would be 9.6 per 100 person-years, almost three times greater than the traditionally calculated figure (e.g. based on underlying cause of death). Finally, if a diabetic death were taken to mean any death in an individual with diabetes, the overall diabetes mortality rate would be 25.8 per 100 person-years. A figure over 2.5 times greater than the all-death certificate rate and almost 8 times higher than the rate based on the underlying cause of death. This pattern was also demonstrated in Ochi JW 1985 study.

Table 9.2 Estimation of diabetes mortality rates using three different definitions.

	Mortality based on diabetes as underlying cause of death		Mortality based on diabetes as underlying or contributory cause		Mortality based on all known diabetes	
Sex and age group	Number	Rate*	Number	Rate*	Number	Rate*
Men						
<24					1	1.19
24-35			2	0.80	5	2.00
35-44	4	1.35	10	3.37	20	6.73
45-54	9	1.91	22	4.67	55	11.68
55-64	13	1.69	60	7.82	161	20.99
65-74	19	2.96	87	13.57	270	42.12
75-84	28	10.49	69	25.84	214	80.15
85+	4	10.00	12	30.00	36	90.00
Subtotal	77	2.73	342	9.30	762	27.05
Women						
<24	1	0.99				
24-34	3	1.03	2	0.69	4	1.38
35-44	10	3.69	4	1.48	10	3.69
45-54	13	3.96	17	5.18	30	9.15
55-64	31	5.56	32	5.73	82	14.70
65-74	25	4.58	81	14.84	202	37.00
75-84	10	3.76	83	31.20	204	76.69
85+	10	12.20	26	31.71	67	81.71
Subtotal	103	4.22	245	10.03	587	25.65

* Mortality rate (per 100 person-years) was computed on a person-year basis by dividing number of deaths by sum of individual follow-up years in each sex and age group.

In conclusion, mortality from diabetes is eight times greater than the estimate obtained on the basis of data from death certificates, which gives only the underlying cause of death with DM mentioned. This problem can be partially overcome by inclusion in the analysis all causes that mentioned DM on the certificate. To estimate the possible impact of diabetes on mortality more accurately, it is important to ascertain all deceased patients in the population with diabetes. Linking all diabetic records with the corresponding records in the NHSCR is the ideal scenario.

9.3 A comparison of the mortality rate in diabetic patients at St Thomas' Hospital with the general population in England & Wales.

Although crude mortality rates have limited value for evaluators, mortality data becomes more interesting once they are standardised (usually by years). The Standardised Mortality Ratios (SMRs) have been analysed in diabetic patients at St Thomas' Hospital from 1980 to 1994 based on annual death rates in England and Wales (Figure 9.2). The methodology for calculating SMRs is described in Section 2 ("Methodology"). This study intends to answer how the mortality rate in our diabetes population is related to year of death, age at death, sex and cause of death by comparing to national figures for England and Wales.

Observed deaths	**1980**	**81**	**82**	**83**	**84**	**85**	**86**	**87**	**88**	**89**	**90**	**91**	**92**	**93**	**94**
Male	22	19	28	33	27	27	36	46	44	63	56	80	82	91	90
Female	21	18	21	22	21	42	28	33	34	53	41	49	44	77	83
Total (a)	43	37	49	55	48	69	64	79	78	116	97	129	126	168	173
Expected deaths															
Male	17	13	21	24	28	33	37	41	45	50	54	57	59	65	70
Female	12	14	16	18	20	23	24	26	29	32	33	36	38	43	46
Total (b)	30	27	38	42	48	56	61	67	74	82	87	94	98	108	117
	1980	**81**	**82**	**83**	**84**	**85**	**86**	**87**	**88**	**89**	**90**	**91**	**92**	**93**	**94**
SMR (a)/(b)*100	145	137	131	131	100	123	105	117	105	141	112	138	129	156	148

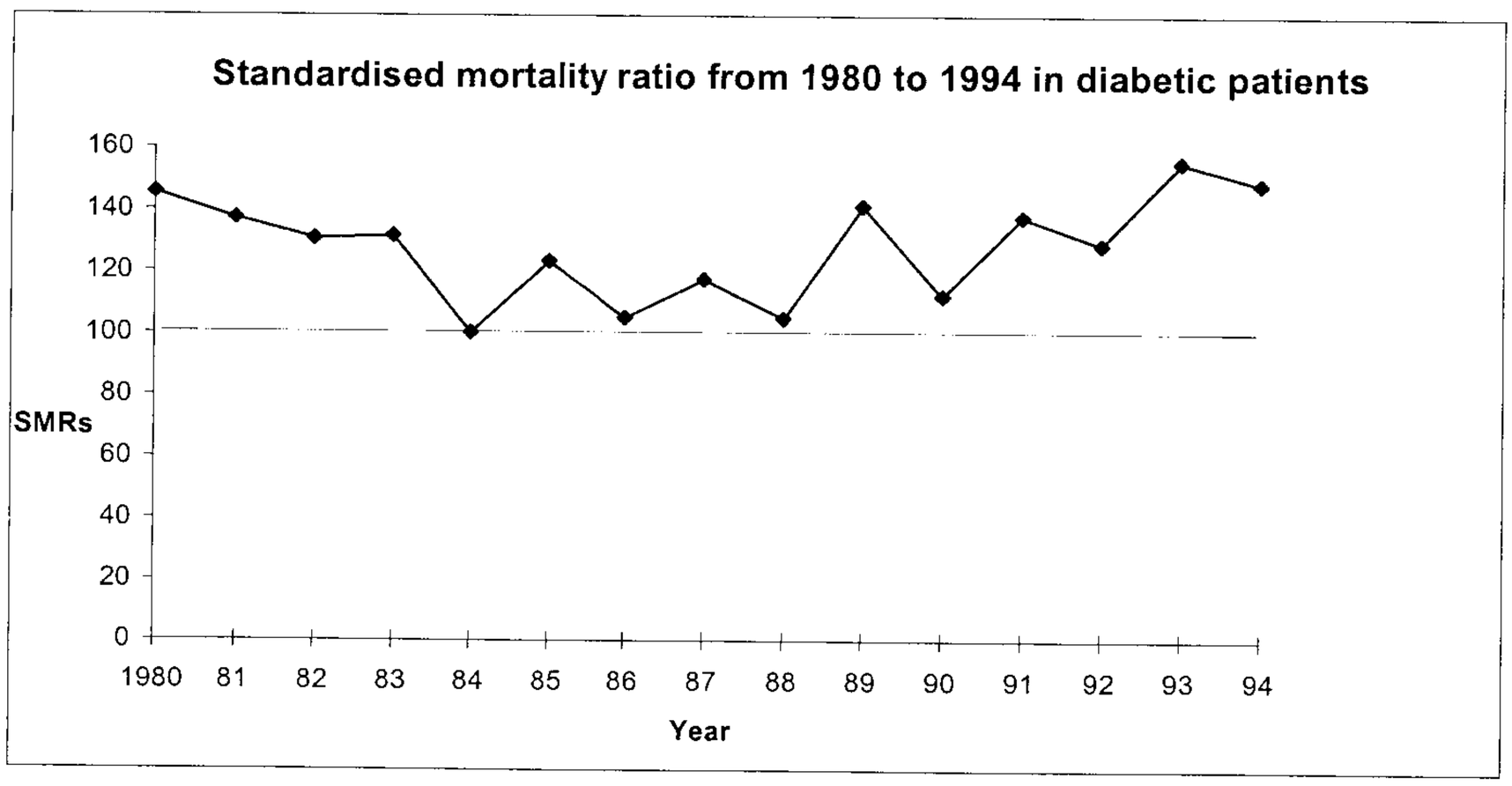


Figure 9.2 All-cause SMRs in diabetic patients at St Thomas' Hospital (1980-1994) by year, based on annual death rate in England and Wales.

Results shown in Figure 9.2 illustrate that the trend of the mortality rate is increasing. Possible reasons of this are discussed later in this Chapter (9.6).

SMRs	1980	81	82	83	84	85	86	87	88	89	90	91	92	93	94
Male	128	141	132	137	97	82	98	112	97	126	104	140	138	140	128
Female	169	133	128	124	105	182	116	126	118	165	124	135	115	178	179

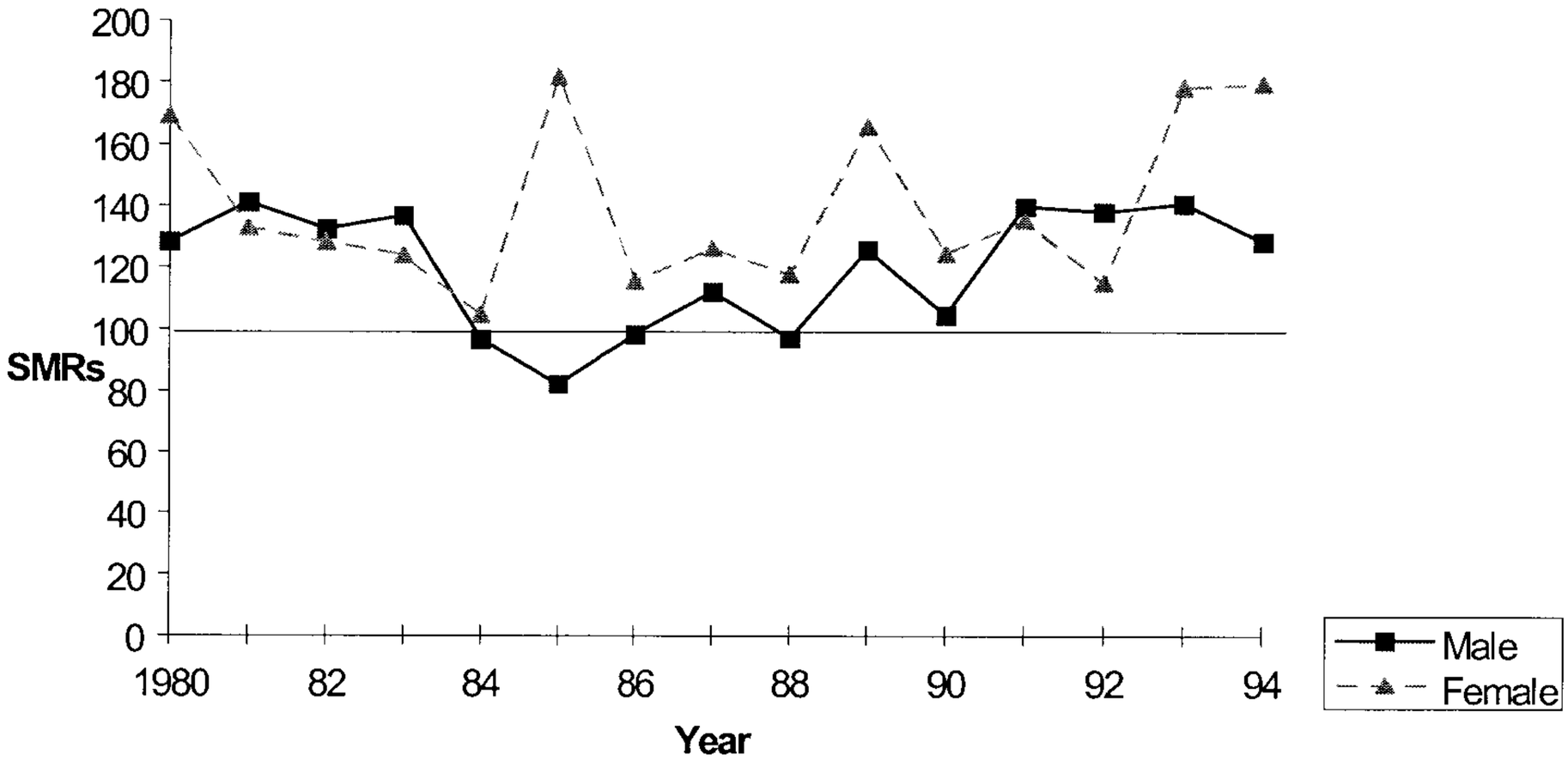
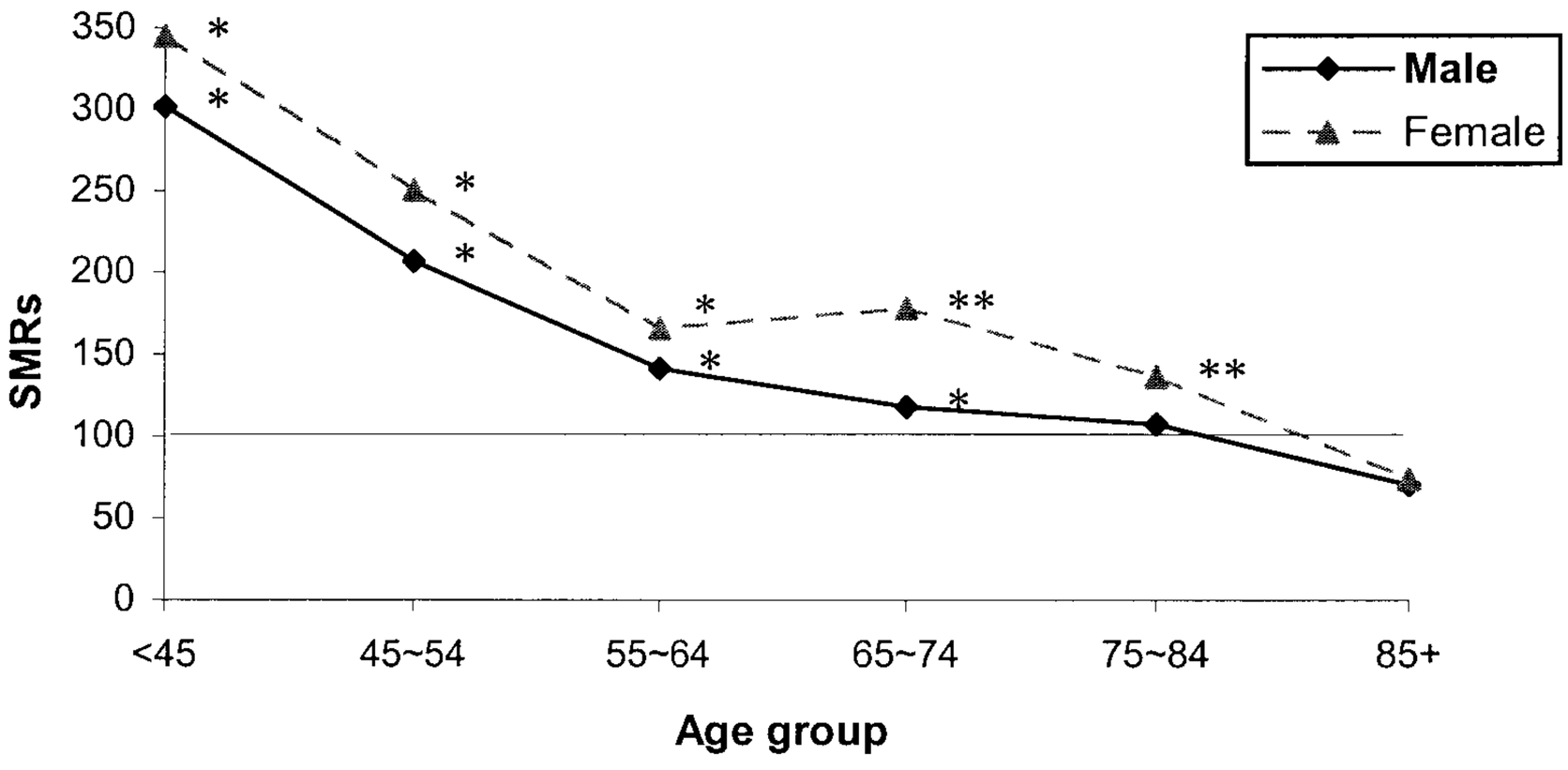


Figure 9.3 All-cause SMRs in male and female diabetic patients at St Thomas' Hospital (1980-1994) by gender and year based on annual death rates in E&W.

The overall SMR was 1.47 [95% CI 1.31~1.62] for women and 1.14 [95% CI 1.04~1.25] for men and shows that the SMRs in female groups were greater than the general population (SMR =100) between 1984 and 1990 (Figure 9.3). Compared to the general population, mortality in the male group was lower (SMRs <100) in the years of 1984, 85, 86 and 88. Mortality in the female groups was much higher in comparison to the general population, especially in the years of 1980, 84, 85, 89, 93 and 94 (SMRs >160). Further research is needed to fully explain these findings.

Age Group	<45	45~54	55~64	65~74	75~84	85+
Male						
SMR	302	207	141	118	107	70
Observed number of deaths	25	53	157	267	209	33
Expected number of deaths	8	26	112	227	196	47
Female						
SMR	345	251	166	178	136	74
Observed number of deaths	14	28	78	199	201	67
Expected number of deaths	4	11	47	112	148	90



Note: p value for SMRs [comparing observed number of deaths to expected number of deaths by using $(O-E)/\sqrt{E}$] * p =<0.05, ** p=<0.01.

Figure 9.4 SMRs in diabetic patients at St Thomas' Hospital, 1980-1994, by age at death for male and female groups, based on annual death rates in England & Wales.

Analysis of SMRs, by age at death, indicates that at all ages the SMRs in female patients were greater than in male patients (Figure 9.4). Mortality in both male and female groups were higher than mortality in the general population (SMRs>100) in age<75. SMRs were significantly increased for age <55.

In conclusion, the all-cause mortality for both males and females in the diabetic population attending St Thomas' Hospital was higher than the general population in England and Wales, particularly in young people. The overall SMRs for females was higher than in males. An excess of mortality existing in people <age 55 yrs may explain this. Moreover there was a trend for increasing SMRs since 1985,

although this appears different in male and female groups. Reasons for this remain unknown, but could relate to the re-allocation of NHS resources away from London in the late 1980's. The association between excess mortality in the diabetic population and the specific causes of death will be fully analysed in the following section.

9.4 Analysis of the specific causes of death in diabetic patients at St Thomas' Hospital, by age and gender

The linkage between "Diabeta" and the NHSCR has enabled us to analyse the causes of deaths in diabetic patients according to information provided on death certificates. In this section, the distribution of the numbers of deaths in the cohort of patients in relation to specific causes of death will be illustrated. The association between such distribution and the age of death will also be discussed.

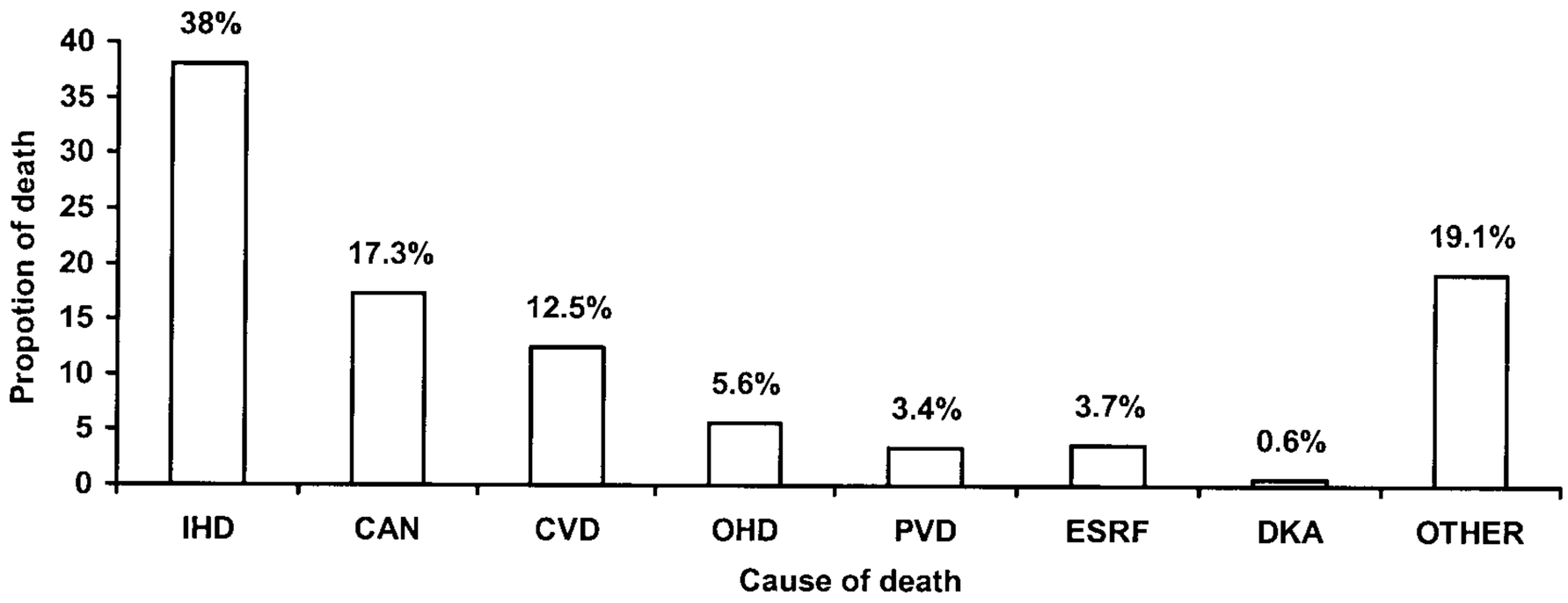
The underlying cause of death (as coded by ICD-9) was used in this study. As a result, completion of death certificates for diabetic individuals was poor (because of a lack of uniform practice amongst medical professionals), the role of diabetes in specific causes of death may need to be redefined (Knuiman MW et al 1991). As discussed earlier, 'Diabetes Mellitus' was unlikely to be mentioned on death certificates. For those, whose underlying cause of death were 'Diabetes Mellitus' (ICD_9 code 250 (diabetes)), the subjects was re-classified where possible to other categories according to the main contributory cause of death by one clinician (Dr David Coppini). This was done to investigate associations between the cause of death and diabetes.

Figure 9.5 Proportion of mortality in diabetic patients at St Thomas' Hospital, distributed by specific causes of death.

Underlying Cause of Death	Number	
Cancer	289	
IHD/MI	635	
Peripheral Vascular Disease	54	
Other heart disease	94	
Cerebrovascular Disease	208	
End Stage Renal Failure	61	
Diabetes Ketoacidosis	10	
Others	319	Details
	1670	

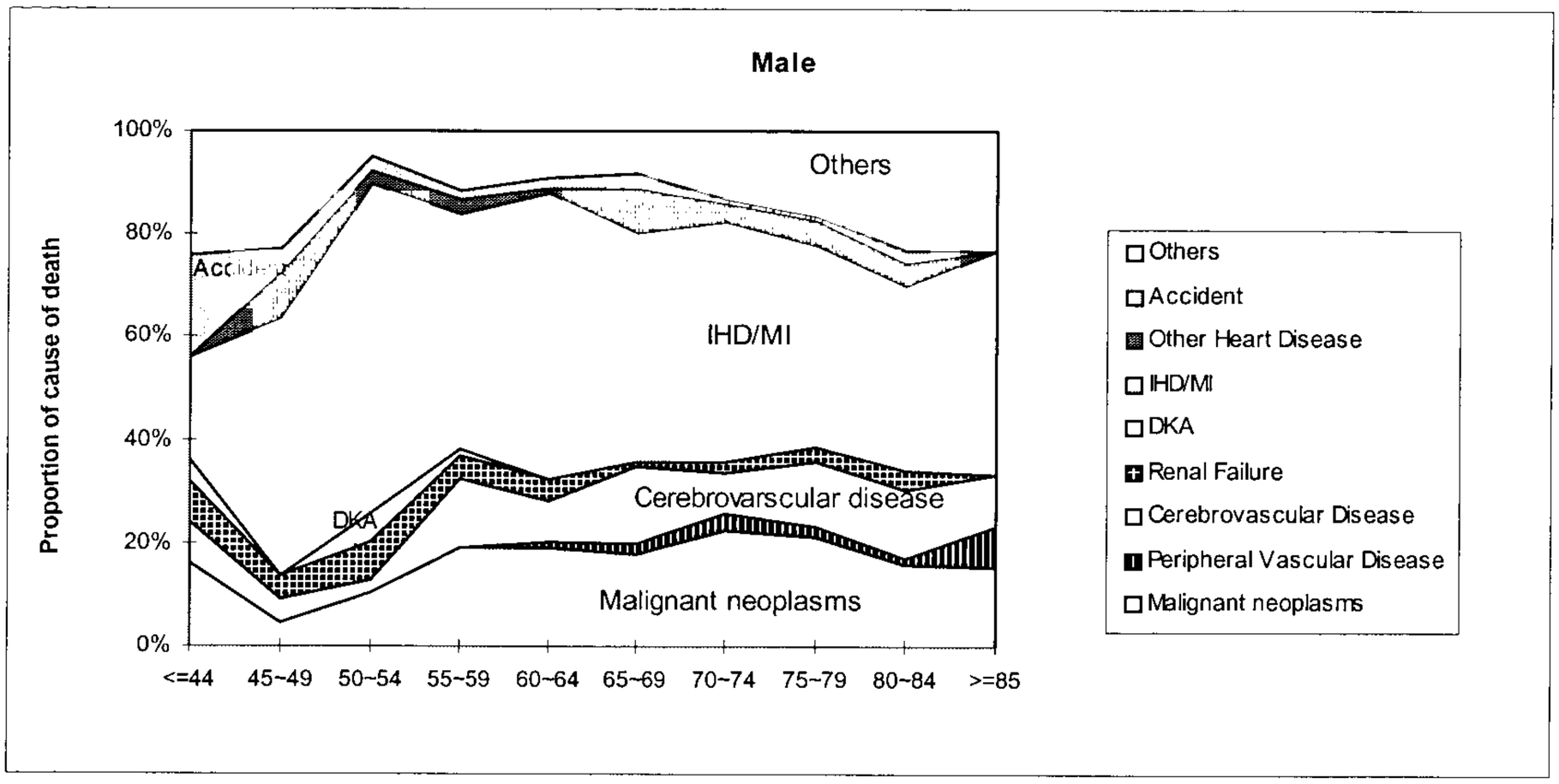
Details of Others (not related to diabetes)

Infection	17
Diseases of Endocrine, Nutritional and Metabolic	14
Pneumonia and Influenza	93
Chronic Obstructive Pulmonary Disease	29
Diseases of the Digestive System	55
Mental Disorders	22
Infection of Kidney	18
Congenital anomalies	7
Uncertain	11
Accident	32
Others	21



* IHD-Ischaemic Heart Disease; CAN-Cancer; CVD-Cerebrovascular disease; OHD-other hearth disease; PVD-Peripheral vascular disease; ESRF-End Stage Renal Failure, DKA-Diabetes Ketoacidosis.

Vascular disease accounted for 897 deaths (53.7%). 635 diabetic patients (38%) died of ischaemic heart disease (ICD 410-414) or myocardial infarction (ICD 430-438), 17.3% patients died of cancer and 12.5% patients died of cerebrovascular diseases (Figure 9.5). This shows that over 50% of patients died of vascular disease, confirming the observations in another study (Wang SL 1996).



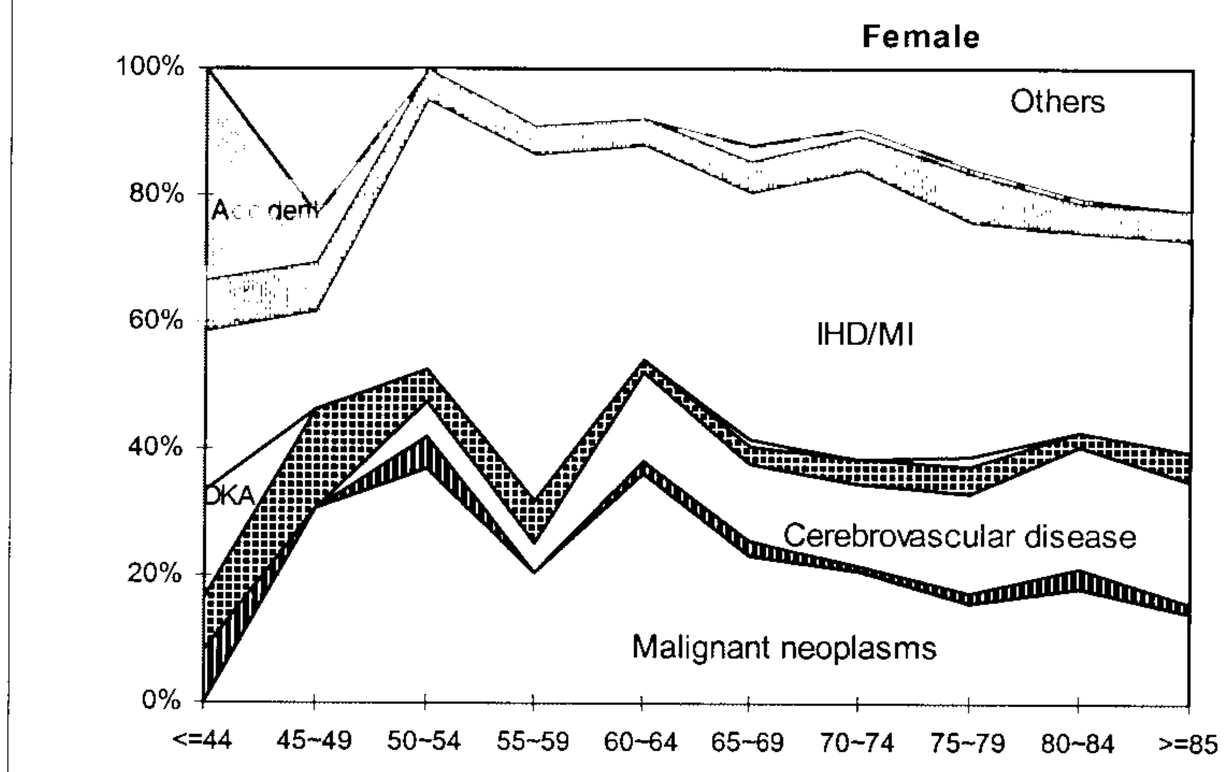


Figure 9.6 Proportion of diabetic patients showing specific causes of deaths, by age at death and gender.

Figure 9.6 shows that for age >50 yrs, the main causes of death in both male and female patients were ischaemic heart disease (IHD), myocardial infarction (MI), cerebrovascular disease (CVD) and cancer (CAN). For age <50 yrs, the causes of death were more likely to be accidental (ICD E800-E999) (e.g. diabetic coma), renal failure, ischaemic heart disease/myocardial infarction and cancer. Details of the causes of death for the age group <50 yrs. are shown in Table 9.3. These findings should be viewed as tentative, because of the relatively small numbers of

deaths in young patients (n=90)(age <50 yrs.). A further study based on a large cohort of patients with age <50 yrs. is required.

Table 9.3 The cause of death in patients' age <50 yrs.

	Number of patients		
Underlying Cause of Death (ICD-9 code)	**Male***	**Female***	**Total***
Ischaemic heart disease (410-414)	21 (36.2)	8 (25.0)	29 (27.8)
Other health diseases (747 or 421-425))	1 (1.7)	2 (6.2)	3 (3.3)
Hypertension (402)	1 (1.7)	2 (6.2)	3 (3.3)
Cancer (140-239)	6 (10.3)	4 (12.5)	10 (11.1)
Cerebrovascular disease (430-438)	3 (5.2)	1 (3.1)	4 (4.4)
Renal failure (585-586)	3 (5.2)	3 (9.3)	6 (6.7)
Peripheral vascular disease (443)	-	1 (3.1)	1 (1.1)
Disease in pancreas (577)	2 (3.4)	1 (3.1)	3 (3.3)
Diabetes ketoacidosis (2501-2502)	1 (1.7)	2 (6.2)	3 (3.3)
Liver disease (571-572)	3 (5.2)	-	3 (3.3)
Infections in lung and bronchitis (485-519)	3 (5.2)	1 (3.1)	4 (4.4)
Infectious diseases	3 (5.2)	1 (3.1)	4 (4.4)
Uncertain death (798-799)	2 (3.4)	2 (6.2)	4 (4.4)
Accidental death#	6 (10.3)	2 (6.2)	8 (8.8)
Suicide (E9530)	-	2 (6.2)	2 (2.2)
Others	3 (5.2)	-	3 (3.3)
Total	58	32	90

* - n (%)

- accidental deaths include road accident, head injury, asphyxia, poisoning by analgesic drugs, burns and hypothermia.

In conclusion, the largest proportion of deaths in the cohort of patients was related to vascular disease. The distribution of mortality was various with age at death. A comparative analysis of mortality related to vascular disease in diabetic subjects to that in the general population in England and Wales is presented in the following section.

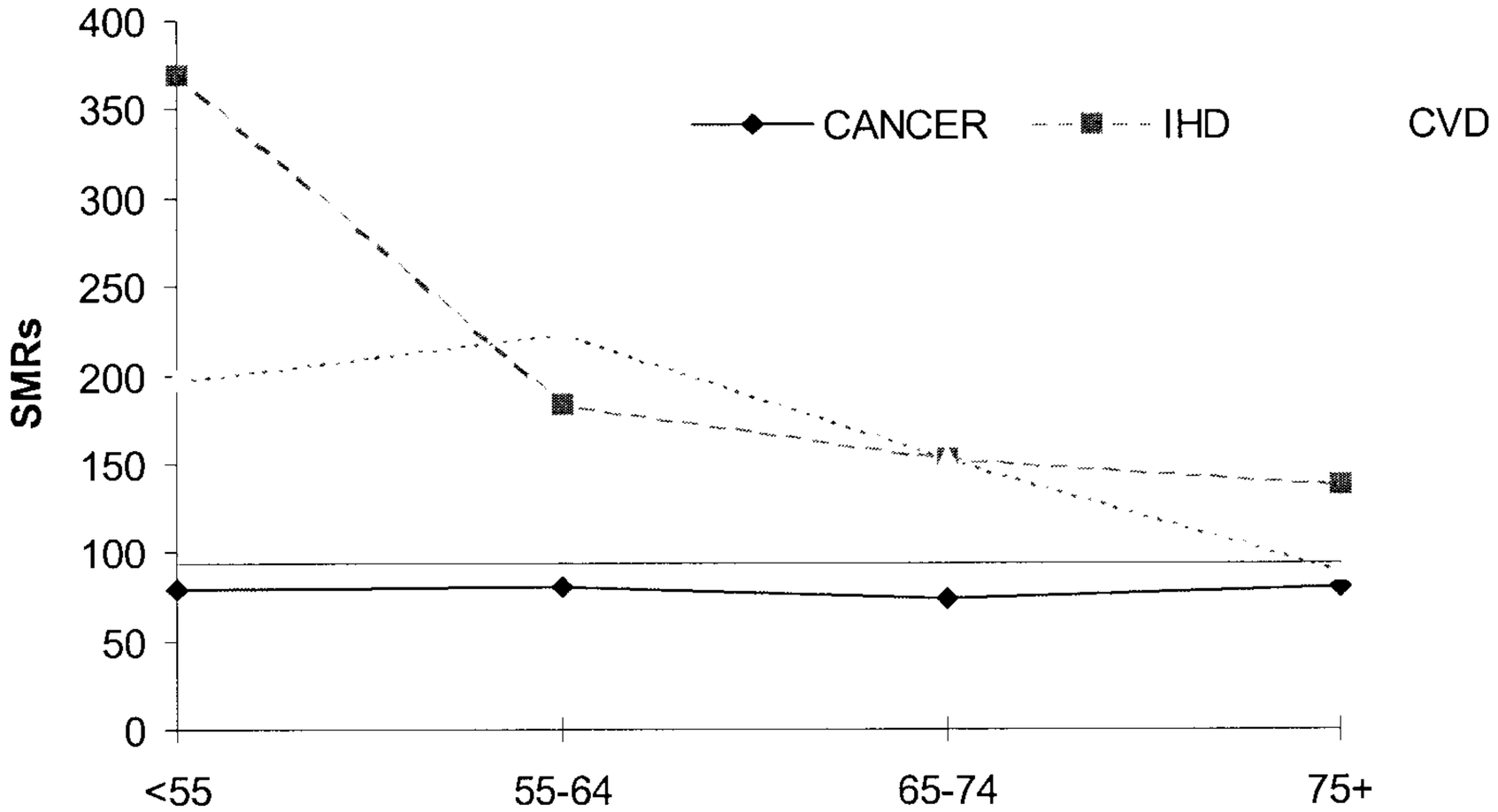
9.5 Analysis of SMRs in specific causes of death by age and gender in the cohort of diabetic patients at St Thomas' Hospital.

Since 1979, diabetes has been the seventh leading cause of death in the United States (Kleinman JC et al 1988). Diabetic patients are also at increased risk of death from other causes, particularly ischaemic heart disease. Previous studies have commonly focused on the role of diabetes as a risk factor for ischaemic heart disease. One seemingly inconsistent finding among these studies relates to whether there is a gender difference in the association between diabetes and ischaemic heart disease death (Kleinman JC et al 1988). In this study, we compare the mortality experience in a cohort of patients from the updated "Diabeta" database, with that in the general population in England and Wales, by gender and cause of death.

The SMRs for the specific causes (ischaemic heart disease, cerebrovascular disease and cancer) of death in diabetic patients at St Thomas between 1980 and 1994 are shown in Figure 9.7.

S.M.Rs by age		<55	*55-64*	*64-75*	*75+*
CANCER		79	80	73	81
IHD	male	369	183	153	138
CVD		196	223	153	90
CANCER		89	105	114	100
IHD	female	884	312	250	148
CVD		327	295	202	101

Male



Female

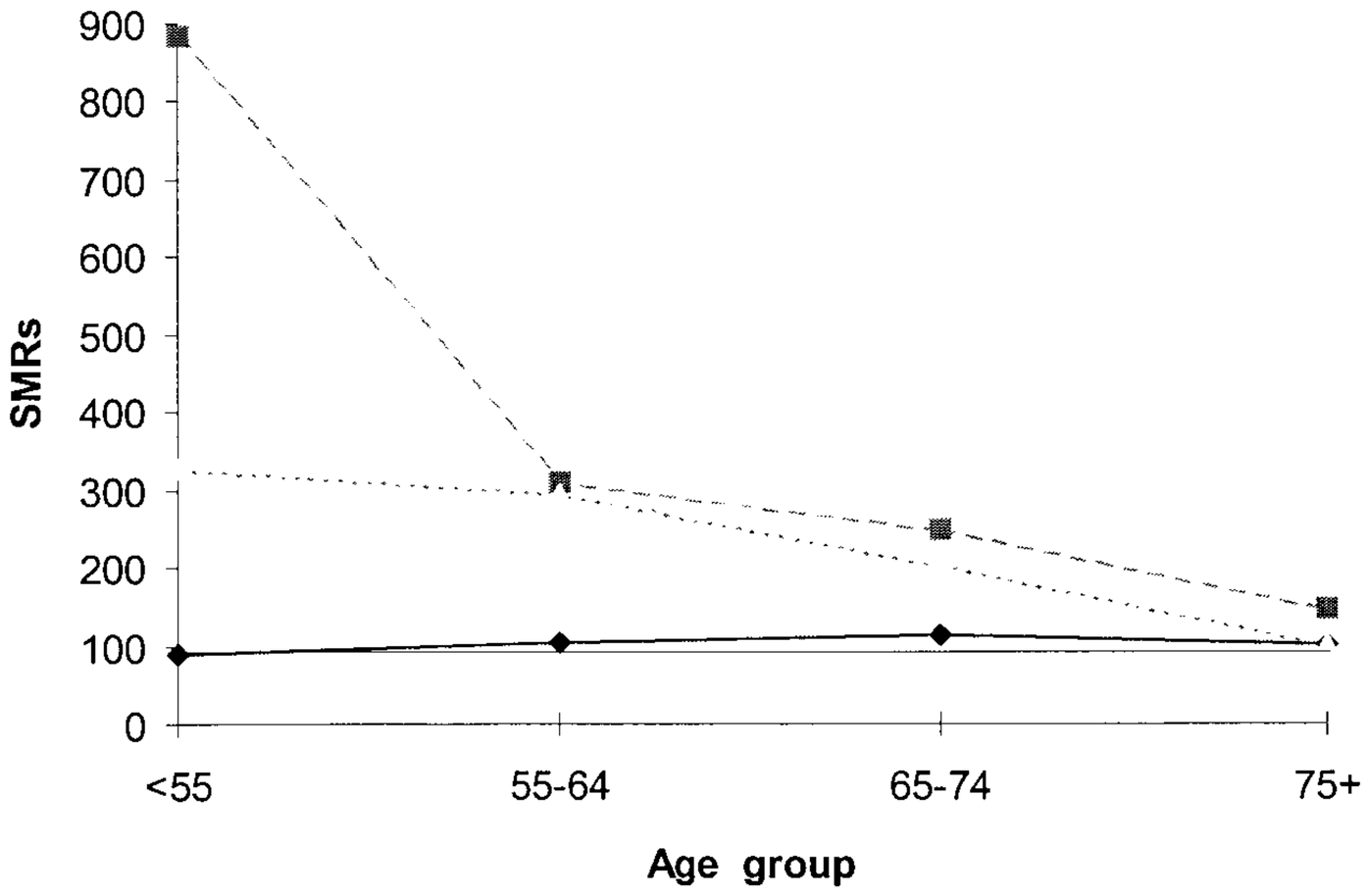


Figure 9.7 SMRs in specific causes of death in diabetic patients (1980-1994) at St Thomas' Hospital by age at death and gender.

Figure 9.7, illustrates that the SMRs for ischaemic heart disease and cerebrovascular disease were greater than 100 for both genders at most age groups. This implies that mortality for IHD and CVD in diabetic patients was higher than that in the general population of England & Wales. In comparison to the general population, the SMR was 2.11 [95% CI 1.65~2.55] for ischaemic heart disease and 2.77 [95% CI 2.24~4.96] for cerebrovascular disease in patients <65 years. For those aged <55, mortality related to IHD was extremely high [SMRs M: 3.69 (2.53~4.85) F: 8.84 (3.83~13.84)]. Another study also demonstrated a similar result where many more deaths amongst diabetic patients were attributed to diseases of the circulation than would normally be expected in the general population (Knuiman MW et al 1992). In conclusion, diabetes accelerates the development of arteriosclerosis in diabetic subjects (Knuiman MW et al 1992). There was no significant difference in cancer related mortality in diabetic patients compared to that in the general population. The SMRs of cancer in the male group were lower than 100 in all age groups. However, in the female group, SMRs were slightly higher than 100. This was recognised in a study by Wong JSK et al 1991, but the cause remains unexplained.

The excess mortality related to other causes of death was not commented upon since the number of deaths was small. These results largely confirm the previous studies (Wong JSK, Aberdeen 1991; Fuller JH, Middlesex, 1983), although results seem to vary between different countries (Ochi JW, Minnesota 1985; Dorman JS, Pennsylvania, US 1984; Kleinman JC, Hyattsville 1984; Sasaki A, Japanese 1981; Green A, Denmark 1980). These studies report a much higher mortality in males than females, as compared to the general population. Geographic, nutrition, life style or cultural factors may be associated with increased mortality in the diabetic population, and this may explain the different mortality rate in different countries. Further discussions will continue in "Section Four".

9.6 Estimation of mortality in diabetes from 1994 to 2010

In order to predict mortality in the future, the author estimated SMRs of diabetes in year 2000, 2005 and 2010 at St Thomas' Hospital by using the mortality data (such as SMRs) in "Diabeta" from 1980 to 1994 (shown in Figure 9.2 in Chapter 9.3). One possible interpretation of Figure 9.2 is that the nadir of SMR in 1984 may be due to the reorganisation of London Health Services. The data was analysed according to where SMR has been fitted to a linear model for the intervals 1980-1984 and 1984-1994 (Figure 9.8).

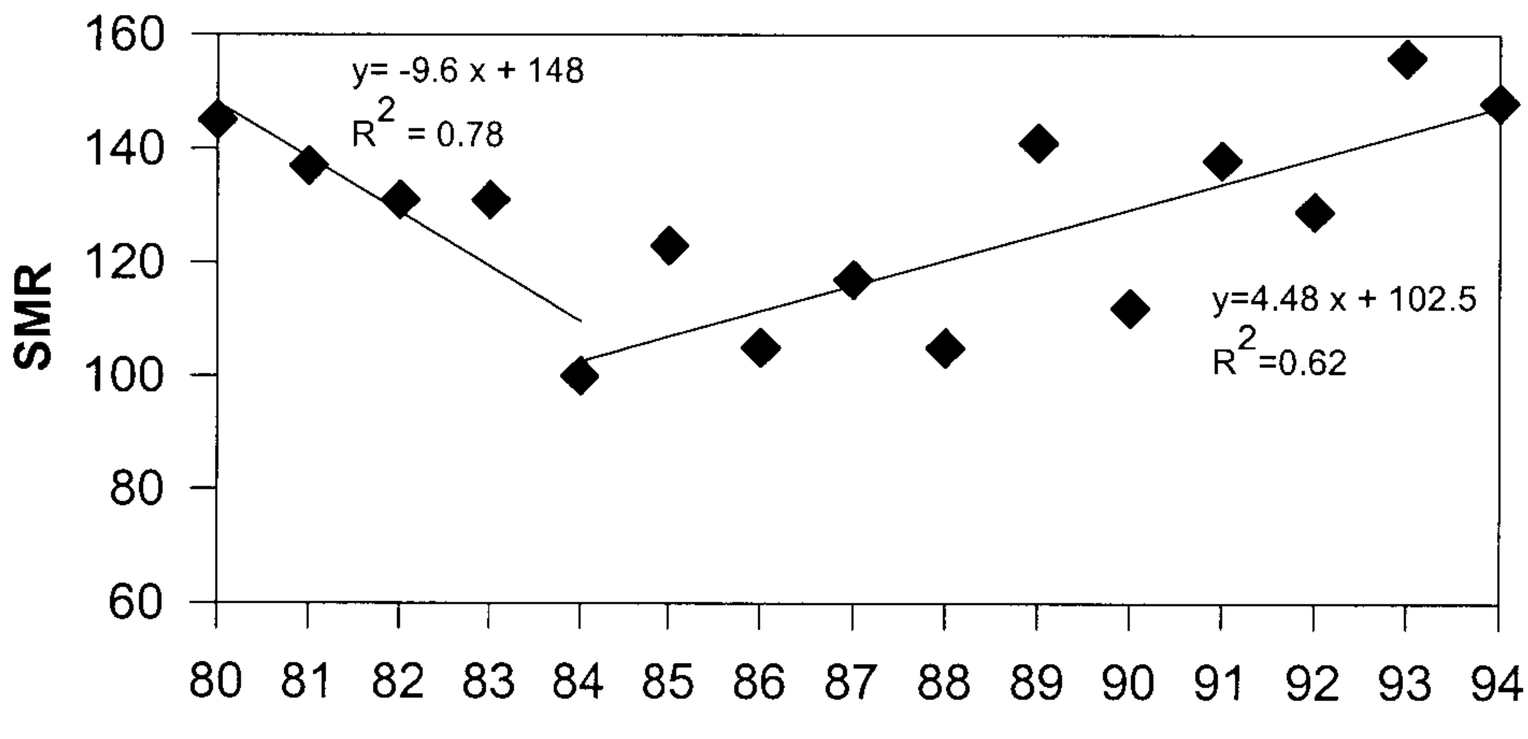


Figure 9.8 Linear-regression in overall SMRs in diabetic patients at St Thomas' Hospital for 1980-1984 and 1984 -1994.

The SMRs in diabetes at St Thomas' Hospital from 1980 to 1984 have decreased. If the SMRs continue to decrease following the trend between 1980 and 1984, the mortality rate in year 2010 in diabetic population at St Thomas' Hospital could be more reduced and comparatively less than the overall mortality in England and Wales.

However if the prediction is based on the trend of SMRs in the recent ten years (such as between 1984 and 1994), the mortality in 2000, 2005 and 2010 will be dramatically increased.

Table 9.4 shows that by the year 2010 the mortality in the diabetic population at St Thomas' Hospital will be almost twice as high as the overall mortality in England and Wales.

Table 9.4 Predicting SMRs for diabetes in "Diabeta", based on mortality data between 1984 and 1994.

Year	**1994**	**2000**	**2005**	**2010**
Overall SMR	148	192	214	237

This study has shown the changes of the trend of mortality over years and may relate to an increase or decrease in premature deaths in our diabetes population. There may be an explanation for this observation. Between 1973 and 1984, our diabetic clinic had succeeded in improving services to people with diabetes living in the Lambeth Southwark and Lewisham area. The number of trained staff working in the speciality was increased and many innovations in health care were made such as: Diabetes Specialist Nurses, Specialist Dieticians, Community Diabetes Liaison Nurses, Home Blood Glucose Monitoring, Low-Dose Insulin Infusions for diabetic emergencies and routine measurement of HbA1c. These service improvements most likely resulted in the falling SMR.

Since the reorganisation of the NHS under the government of that time and the implementation of the RAWP (Re-Allocation Working Party) recommendations (re-allocation of money away from 'over-provided' London to the 'under-provided Shires'), the number of Community Diabetes Liaison Nurses working in the Lambeth area have been reduced to zero whilst the population of people with diabetes has increased. The services for people with diabetes in inner London are indeed at breaking point and can no longer adequately cope.

The reallocation of NHS resources away from London the late 80s is only partially responsible for an increase in diabetes mortality. Other reasons may be that now GPs are increasingly trying to look after patients with diabetes, and hospital case mix has changed with the advent of more advanced disease that has resulted from an increase of mortality in diabetes.

The author's study shows an alarming trend to an increase of excess mortality in the diabetes population. There is an urgent need to develop an appropriate monitoring system for assuring quality of diabetes services via information technology. This study could provide a useful guide for targeting primary & secondary prevention programs and may also serve as an impetus to reduce the global burden of diabetes and excess mortality predicted in this thesis for the next century.

The author's findings, in term of SMRs, were also compared with other areas. In Figure 9.10, the overall SMRs for diabetes in the period 1995 to 1997 in national/local areas given by Public Health Common Data Set 1998 from the Department of Health, were compared with the SMRs (from 1980 to 1994) for diabetic patients registered in "Diabeta" at St Thomas' Hospital. All SMRs are estimated on the basis of the annual death rate in England and Wales within the corresponding years. The SMRs in national/local areas for selected causes of death (e.g. ICD9 250 Diabetes Mellitus) were assessed based on death certificates alone, but not in St Thomas' Hospital (here it was based on the entire diabetes subjects registered on "Diabeta"). The confidence intervals - lower limit (SMRLL) and upper limit (SMRUL) were calculated on 95% level. A SMR greater than 100 indicates excess mortality of diabetes in a given area/health care setting as compared with the general population in England and Wales.

Areas	Period time	OBS *	SMR	SMRLL	SMRUL
England and Wales (E&W)	1995-1997	8424	100		
England	1995-1997	7870	99	97	102
North Thames	1995-1997	1020	100	94	107
South Thames	1995-1997	1029	91	112	133
Inner London	1995-1997	570	122	133	135
Lambeth, Southwark & Lewisham (LSL)	1995-1997	101	111	90	135
Lambeth	1995-1997	41	136	98	185
Southwark	1995-1997	30	101	68	145
Lewisham	1995-1997	30	96	65	137
St Thomas' Hospital (STH)	1980-1994	1331	147	131	162

* OBS: observed number of deaths

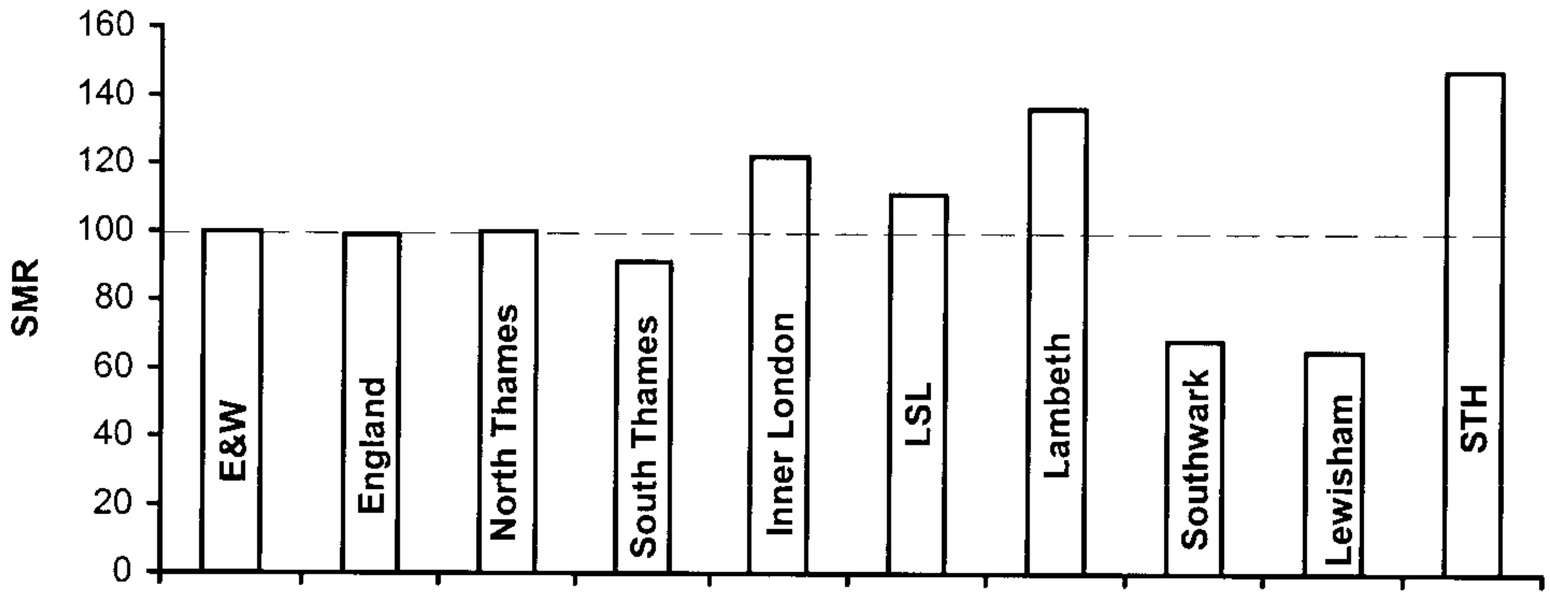


Figure 9.10 Comparison of overall SMRs between St Thomas' Hospital and other areas.

The SMR for diabetes patients at St Thomas' Hospital proved to be the highest, when compared to other areas. Firstly, this could be due to mortality being assessed on the basis of all known diabetic subjects registered on "Diabeta", rather than according to the words 'Diabetes Mellitus' mentioned on death certificates (like other studies). Secondly, it is assumed that the more ill the patients are, the more likely they are to be treated in the hospital. The SMR for diabetes at St Thomas' Hospital was very similar to the SMR for diabetes in the Lambeth area, and this may due to the fact that most diabetic patients (75%) registered with St Thomas' Hospital lived in the Lambeth area. The true figure of SMRs in a local area (e.g. LSL) would be higher if the estimate was based on all diabetic subjects, and not solely based on death certificates mentioning 'Diabetes Mellitus'. To have

a meaningful comparison of mortality, linking diabetic subjects' records with death records is essential.

9.7 Conclusions

Underestimation of diabetes-related deaths is a common problem for mortality studies in the diabetic population. It can, however, be overcome by linking diabetes patient records with corresponding records in the NHSCR. Linkage between "Diabeta" and the NHSCR has enabled us to access the causes of death on death certificates and to accurately assess mortality in our diabetic patients. This linkage allows an accurate on-going estimate of mortality in our diabetic patients.

As over 50% patients registered on "Diabeta" are from the local health authority (LSL) boundary area, this linkage should be useful in helping to focus the appropriate resources to diabetic patients in this community.

Comparison of the diabetes related mortality in St Thomas' Hospital with other health care providers (e.g. hospital, GP, community services) would be possible, only if such providers produced reliable information on deceased patients.

Chapter 10

Identification of risk factors related to mortality within diabetic patients

In previous chapters the author has demonstrated that diabetic subjects have a greater age-specific mortality than the general population, as confirmed by other studies (Panzram G 1987, Fuller JH et al 1983, Shenfield GM et al 1979, Waugh NR 1989, Wong JSK et al 1991). This propensity for excess mortality still remains unexplained. Since 'Diabetes Mellitus' is under reported on the death certificates, few prospective observational studies have been implemented to investigate the determinants of mortality in diabetic subjects through different methodologies. The aim of this study is to determine baseline predictors of mortality in a defined cohort of patients attending our clinic between 1982-85, based on the updated "Diabeta" database.

10.1 A cohort of long-term follow-up patients used in the study

A cohort of all 1000 consecutive patients first attending the diabetes clinic at St Thomas' Hospital between January 1982 and September 1985 was selected from "Diabeta". The main aim of the study was to obtain information on the natural history of end-stage neuropathy (ulcers and amputations) over 10 years by reviewing patients in 1995. The author, assisted by Dr David Coppini, studied the risk factors that may be associated with excess morbidity and premature mortality in our diabetic population.

Of 426 (42.6%) patients reviewed personally in the clinic by Dr Coppini, half of these had not attended the diabetes clinic for at least 2 years due to patient re-location. However their new location was obtained through updated FHSA data provided by the records linkage between "Diabeta" and the NHSCR. A standard letter was sent to GPs asking for their co-operation to recruit patients for this study. 50 patients (5%) went abroad and 120 patients' records (12%) could not be traced in the NHSCR systems. 147 (14.7%) patients who, for various reasons (mainly long distance travel or disability), could not be reviewed clinically, were sent a standard questionnaire on foot ulcers or amputation. 248 patients from the cohort died between 1982 and 1995, and death certificates were provided. Data on any lower

extremity events was also collected, where possible, from notes (usually microfilmed) in various hospitals or long-stay institutions. This proved to be useful, in that lower extremity amputations were detected in this way in 2.4% of deceased patients, as none had been previously recorded on the death certificate itself. A breakdown of the cohort at review is illustrated in figure 10.1.

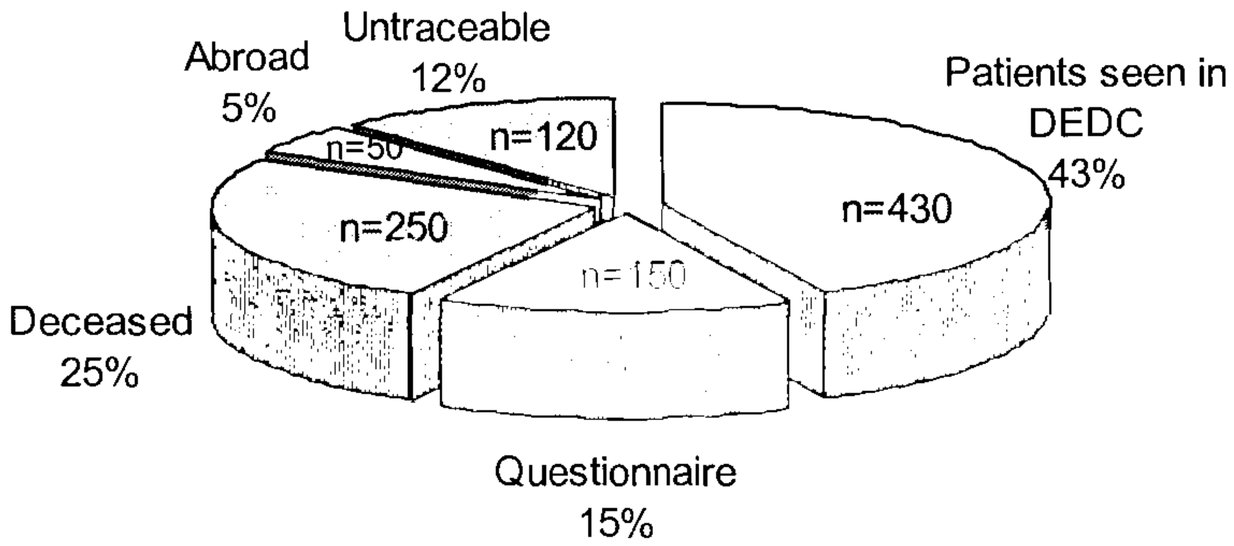


Figure 10.1. Follow-up of patients in cohort (n=1000)

The record linkage between "Diabeta" and the NHSCR has enabled a mean 12 year prospective analysis on a selected cohort, and has also provided updated demographic and clinical information on >80% patients (Coppini DV et al 1998). This is one of the longest prospective studies, on a well-defined cohort of diabetic individuals in the UK. This cohort of patients has enabled the author to carry out further studies related to risk factors (clinical and socio-economic) for increased mortality and morbidity in the diabetic population.

10.2 Use of an Accelerated Failure Time model (ACF) to analyse the risk factors related to increased mortality in diabetic patients

An Accelerated Failure Time (ACF) model is used to analyse risk factors related to increased mortality in diabetic patients. The statistical methods used for the analysis are described under "Methodology" (Chapter 6). The ACF was chosen as any variable that would be expected to 'accelerate' a patient's risk of death. The key variables measured were:

- Type of diabetes (Type 1 and Type 2)
- Gender
- Race (Afro-Caribbean, Asian, Caucasian)
- Body Mass Index (BMI)
- Duration of diabetes (years)
- HbA1
- Toescore*
- Retinopathy
- Proteinuria
- Alcohol
- Smoking

* Toescores are Vibration Perception Threshold (VPT) measurements at the great toe (Biothesiomether - Biomedical Instrument, Newbury, Ohio, USA), which are then converted to standard score by log-transformation adjusted for age, (Bloom S et al 1988). Correction of VPT for age is a useful and reliable way of identifying patients with peripheral neuropathy (Coppini DV et al 1996).

Systolic and diastolic blood pressure were excluded from the model due to some missing values at baseline.

From the data presented, the time to failure is taken as the age of the patient's death, or when they were last know to be alive (censor observations). To identify the prognostic variables that significantly influence mortality, a stepwise procedure was used to accept or reject variables (at the 5% significance level) from the ACF model. Due to missing values (55 moved abroad and 11 were incorrectly diagnosed as diabetic) and missing values for some of the key variables involved in the analysis (mainly BMI and Smoking), data on 191 (19.4%) patients had to be

disregarded, leaving 794 of which 201 (25.3%) had died. Of the 180 diabetic patients removed, 37(20.6%) had died. Consequently, the proportion of deaths amongst omitted patients to those used in the model, are similar. Furthermore, the distributions of prognostic variables in the two groups were similar to those variables where partial data was available. Due to these similarities, the removal of the 180 patients should not 'bias' the analysis. The prognostic variables of interest at the baseline visit (1982-85) and their coding for analysis in the statistical package SAS are given in Table 10.1. Also given are the percentage of patients falling into each level for prognostic variables, for survivors and those who died.

Table 10.1 Baseline variables and characteristics considered in the model (together with SAS® coding).

Prognostic Variable	Levels	Coding for SAS	Survivors(%) n = 593	Dead (%) n = 201
Discrete Variables				
Alcohol	User	U	60.7	58.4
	Non-User	NU	39.3	41.6
Proteinuria	Absent	0	91.8	79.1
	Present	1	8.2	20.9
Race•	Afro-Caribbean	Afro	20.5	6.5
	Asian	Asian	7.0	3.0
	Caucasian	Cau	72.5	90.5
Retinopathy	Absent	0	82.7	82.6
	Present	1	17.3	17.4
Sex	Female	F	43.2	43.8
	Male	M	56.8	56.2
Smoking	Smoker	S	22.6	25.3
	Non-Smoker	NS	77.4	74.7
Type of Diabetes	Type I	1	30.5	9.4
	Type II	2	69.5	90.6
Continuous Variables				
BMI			28.1 ± 5.1	27.4 ± 4.9
Duration			5.9 ± 8.7	4.4 ± 8.5
HbA1			12.0 ± 3.3	12.5 ± 3.4
Toescore			9.9 ± 1.2	10.2 ± 1.4

Discrete variables summarised as percentages. Continuous variables summary given as mean ± SD.• entered model, but was removed after other variables added

Table 10.2 shows the results of model building, giving the p values at which significant variables entered the model. Also, the acceleration factor ratios, and confidence intervals are given. For derivations, see Table 10.1. The most significant contributor to mortality was the 'toescore', with a lower score providing a protective effect. Of the two binary variables in the model, the most important was the presence of proteinuria, then the type of diabetes with Type 1 diabetic patients clearly having a worse outcome. It is clear that diabetic microvascular complications are strongly related to a reduced survival time. Of the three complications (retinopathy, proteinuria and neuropathy), neuropathy had the highest significance. Race was a significant risk factor when considered independently, but was dropped from the model (because it was no longer significant) when other variables were introduced. However, it has shown in Table 10.3 that the age-sex adjusted mortality rate in Afro-Caribbeans in the cohort is significantly lower than Caucasians (0.72 vs. 2.76 per 100 person-years). Other risk factors studied were glycated haemoglobin (HbA1) at baseline, gender, smoking and alcohol consumption, none were significantly related to increased mortality, and were dropped from the model.

Table 10.2 P-values, acceleration factor ratios with approximate 95% confidence intervals, difference in median survival times (where calculable) for variables found to be significant in model.

Variable	p-value	Acceleration	95% C.I.	Difference in median survival times (years)
Toescore	0.0003	1.2144	1.09 - 1.35	
Proteinuria	0.0020	1.7377	1.22 - 2.47	4.89
Type I Diabetes	0.0249	1.7481	1.07 - 2.85	4.94

Table 10.3 Age-sex adjusted mortality rate by race in the cohort of patients (n=958).

	Caucasian			**Afro-Caribbean**			**Asian**		
Age (yr.)	Number of persons years followed	Number of deaths	Mortality rate	Number of persons years followed	Number of deaths	Expected number of deaths[@]	Number of persons years followed	Number of deaths	Expected number of deaths[@]
Males	n*=420			n=75			n=47		
15-24	111	4	3.60	12	-	0.43	10	-	0.36
25-34	517	0	0	30	-	0	13	-	0
35-44	782	2	0.26	83	-	0.21	40	-	0.10
45-54	860	8	0.93	193	1	1.80	153	1	1.42
55-64	1172	17	1.45	291	3	4.22	250	3	3.63
65-74	1000	42	4.20	229	4	9.62	103	1	4.33
75-84	402	38	9.45	65	1	6.14	30	1	2.84
85+	42	10	23.81	21		5.00	2		0.48
Females	n=311			n=88			n=17		
15-24	237	1	0.42	399	2	0.04	0	-	0
25-34	428	2	0.47	872	-	0.29	5	-	0.02
35-44	347	0	0	839	-	0	9	-	0
45-54	401	4	1.00	1777	-	3.36	23	-	0.23
55-64	684	12	1.75	2987	1	6.49	72	-	1.26
65-74	794	37	4.66	2566	2	9.55	70	-	3.26
75-84	466	29	6.22	1192	-	2.24	20	-	1.24
85+	41	9	21.95	205	-	0.88	0	-	0
Total	8284	215	2.6		14	50.27		6	19.17
SMR			1			0.28 (14/50.27)			0.31 (6/19.17)
Age-sex Adjusted Mortality Rate			2.6			0.72 (0.28×2.6)			0.81 (0.31×2.6)

@: Expected number of deaths among patients who are Afro-Caribbean/Asian are estimated by assuming that their mortality rates were the same as those of Caucasians.

Methodology: * n is number of person in each group

Although various studies have identified risk factors for increased all-cause mortality in diabetic patients, little data is available on the association with diabetic neuropathy. Age-corrected vibration perception thresholds (toescores) provide a simply useful objective measure of neuropathy and are useful predictors of both foot complications (Young MJ et al 1994) and increased mortality in diabetic patients (Weng C et al 1996). Interestingly, higher 'toescores' were more strongly associated with increased mortality than other microvascular complications. Although Vibration Perception Thresholds (VPT, Biothesiometer) is a measure of peripheral sensory neuropathy, the observed increased mortality may be related to subsequent foot ulceration and gangrene (Weng C et al 1996) or concurrent autonomic disturbances, leading to increased cardiovascular mortality. The relationship between mortality and proteinuria is well establish in Type 1 diabetes (Rossing P et al 1996; Viberti GC et al 1982; Borch-Johnsen K et al 1985) and has also been reported in Type 2 diabetes (Jarrett RJ et al 1984). The prevalence of microvascular complications in our cohort was comparable to those in other studies. The Wisconsin study (Klein R et al 1989) showed a prevalence of retinopathy of around 25% (19% in this study) after 5 years of diabetes. The prevalence rates of proteinuria varied from 3% to 16% (Fabre J et al 1982), whereas our cohort was 12.3%. The use of 'toescore' as a measure of neuropathy is unique to this study. A large UK Multicentre Study (Young MJ et al 1994) on prevalence of neuropathy showed a mean raw VPT of 21.1 Volts in diabetic patients, compared with a VPT of 14.9 Volts in this study. This is likely to be related to the longer duration of diabetes in the UK Multicentre Study.

The present study shows that other risk factors such as glycated haemoglobin (p value: 0.577) and smoking (p value: 0.144) were not significantly related to increased mortality. Although glycated haemoglobin at the baseline visit was not related to increased mortality, subsequent glycaemic control may have been a more important determinant on outcome. A similar effect may be seen with other conventional risk factors such as smoking, if for example, the risk of long term rather than baseline cigarette consumption had been considered. Another reason for the non-significant smoking effect may be due to 'under reporting' of tobacco consumption, which can lead to non-significant results (Woodward M et al 1991).

The effect of race on mortality was significant when considered independently, however introduction of the neuropathy variable resulted in the loss of such an effect. This may due to the observation that neuropathy was rare in Afro-Caribbean and higher in Caucasian (Coppini DV et al 1997). In a separate study we have shown that Caucasians are more susceptible to neuropathy than Afro-Caribbean patients (Coppini DV et al 1997). Diabetic subjects still have a reduced life expectancy despite the many potential advances in patient care. We have shown in this study that microvascular diabetic complications are important markers for those at greatest risk and that early diagnosis of diabetes and detection of the complications of diabetes (e.g. neuropathy etc.) is essential for prevention of premature mortality in the diabetes population.

Chapter 11

Identification of social and geographical factors related to increased morbidity and mortality in diabetic patients

Social and economic deprivation was associated with illness in the reports of Black and Townsend. There is also evidence that people living in wealthier parts of a community tend to be healthier (Black D et al 1980; Townsend P et al 1992; Ward JD et al 1994). For many years, the standardised mortality ratios for persons living in industrial towns in the north of England have exceeded the national average (Carstairs V et al 1989). It is also likely that large cities as complex as London have special and unique problems, which makes the delivery of care particularly difficult. Problems including high rates of patient mobility, homelessness, unemployment, specific needs of large ethnic minority groups, drug abuse and AIDS complicate the provision of suitable preventative care for chronic diseases such as diabetes (Sönksen PH et al 1993). The evidence for a significant effect of social deprivation on the outcome of diabetes care in London needs to be clarified.

St Thomas' Hospital is located in the "heart" of London and most of its surrounding area has been defined as a deprived area (LSL Health Authority executive summary 1995/1996). A large proportion of diabetic patients (70%), who are registered in the diabetic clinic at St Thomas' Hospital, live in the Greater London Boundary since 1979 (Newsletter of the Juvenile Diabetes Foundation (UK) 1996). The record linkage between "Diabeta" and the NHSCR is singularly responsible for establishing an accurate mortality study in those diabetic patients. The relationship between social and economic risk factors and morbidity and mortality in diabetic subjects will be demonstrated in this Chapter through linking a geographic information software (MapInfo) with updated patient's data (demographic/clinic) in "Diabeta".

11.1 The cohort of patients used in the study

The cohort of 1000 patients in Chapter 5 described previously was also used in this study. Out of this cohort, patients (n=610) who lived within the Greater London Boundary, were selected for the analysis. Out of these, 186 (30%) patients had died by December 1995, and 332 (54%) were re-examined in the diabetes clinic. 92 (10%) patients were untraceable (or could not be traced in the NHSCR due to incorrect demographic data). The aim of the study was to use postcodes to determine whether the outcome on diabetes care was linked to material deprivation and domicile. The ranking of deprivation was made by using the Jarman Under Privileged Area (UPA) score. The UPA score (widely used index of social deprivation; Jarman 1995) is calculated by combining eight variables derived from the decennial census data at ward level (Jarman B et al 1983; Irving D et al 1983; Morris R et al 1991). The variables were created by a national sample of General Practitioners in the UK who were considered to have an increased workload or pressure on their services. The Department of Health agreed to provide a deprivation allowance to General Practitioners who had patients on their list who lived in areas identified as deprived on the UPA score. The following eight variables were included: elderly people living alone, one-parent families, children < 5 years of age, social class V (unskilled workers), unemployed (as % of economically active population), overcrowded households, people changing house within the last year and those born in the New Commonwealth or Pakistan.

According to a government report, a UPA score >30 is associated with a deprived area, and a score <10 to a prosperous area. A score of 10-30 is related to an intermediate deprived area (Jarman B et al 1983).

A geographical information software system (MapInfo) has enabled us to illustrate and analyse the geographic distribution of a sample of 332 patients at an electorate ward level in the Greater London area, and each ward defined by UPA scores (Figure 11.1-11.2). There are 181 patients (54.5%) from 43 deprived wards and 92 patients (27.7%) from 44 wards which have rankings of deprivation between 10 and 30 (intermediate deprived area) in Greater London. In comparison, the 54 least deprived wards in this study, referred to as prosperous, accounted for 59 patients (17%). The definition of neuropathy was based on Vibration Perception Threshold

(VPT) testing using a Biothesismeter. Patients with a VPT>1.96 SD above, the mean value for age were classed as having neuropathy (Coppini DV et al 1998). The presence of retinopathy was based on any fundal abnormality on direct light ophthalmoscopy. Proteinuria was defined as a single positive albustix test on a morning urine sample.

11.2 Effect of socio-economic factors on outcome of diabetes

By comparing the difference between patients from deprived and prosperous areas in terms of demographics, treatment, diabetic control and clinical outcome status, the following results are found:

Patients living in deprived wards were older at the time of the study, with a shorter duration of diabetes and were also over weight than patients from prosperous wards. (Table 11.1). Patients from deprived wards had more clinic visits than prosperous wards. Although more Afro-Caribbean's live in deprived areas than those in prosperous (36% vs. 5%), however a high proportion of patients from deprived areas were still Caucasian (55% vs. 36%) (Table 11.2). Mean HbA1 for patients from deprived wards was significantly higher than that of patients from prosperous wards (Table 11.1). Insulin treatment was used less commonly in deprived areas ($p< 0.04$) and such patients showed poorer glycaemic control (HbA1 11.2% $\pm$ 2 vs. 8.9% $\pm$ 4 $p<0.01$) (Table 11.4). Patients living in deprived wards were more likely to be smokers than patients from prosperous wards (62.7 vs. 45.30% $p<0.02$). Conversely, patients from prosperous wards were more likely to be significant alcohol consumers (counted as Yes/No) (74.57% vs. 60.22% $p<0.04$) (Table 11.3). The age-adjusted prevalence of microvascular diabetic complications (neuropathy, lower extremity complications (ulcer/amputation) and proteinuria but not retinopathy) was significantly higher in deprived areas (Table 11.6a). Patients from deprived areas were more likely to develop more complications than prosperous areas (Table 11.6b).

As the number of deaths in the cohort is small, the association between mortality and socio-economic factors was identified and based on all diabetes registered (n=5627) on "Diabeta" between 1980 and 1994 (all of whom were residents in the Greater London area). The age-sex adjusted mortality rate in patients from

deprived areas was significantly higher than that in prosperous areas (2.66 vs. 1.91) (Table 11.7). Mortality related to ischaemic heart disease and cancer was also significantly higher in deprived areas, and this was assessed in relation to the specific cause of death (Table 11.8).

Table 11.1. Relationships between diabetic patients' demographic/clinical data and their social deprivation scores categorised.

	Deprived	Intermediate	Prosperous	P value[#]
Number of patients (n)	181	92	59	
Age at 1995*	61.3 ±12.2	61.3 ± 13.4	58.6 ± 13.6	0.014
Duration (yr.)*	14.2 ± 6.5	14.2 ± 5.5	19.6 ± 11.0	0.002
BMI*	29.2 ± 5.3	27.2 ± 6.1	25.7 ± 6.4	0.003
Male sex (%)	55.8	51.1	52.5	0.08
Number of visits*	11.3 ± 11.1	9.9 ± 9.9	8.3 ± 4.1	0.03
HbA1*	10.5 ± 2.6	9.9 ± 2.4	9.1 ± 3.6	0.003

* Data is mean ± SD or otherwise stated;

[#] t-test for mean value and χ^2 for proportion value; p value shows the significant difference between deprived and prosperous groups.

Table 11.2 Relationships between race and social deprivation categories.

	Number of patients		
	deprived*	Prosperous*	P value[#]
Caucasian	100 (55)	51(86)	
Afro-Caribbean	66 (36)	3 (5)	<0.005
Asian	15 (8)	5 (8)	
Total	181	59	

* Data is n (%);

[#] χ^2 test, p value shows the significant difference of proportion of people of each ethnic minority in deprived and prosperous groups.

Table 11.3. Relationships between treatment types and social deprivation categories.

	Number of patients			
	Deprived*	Intermediate*	Prosperous*	P value#
Insulin	70(38)	38(41)	35(60)	0.004
Diet	15(8)	9(9)	2(3)	0.134
Tablet	96(53)	45(49)	22(37)	0.026
Total	181	92	59	

* Data is n (%);

χ^2 trend test, p value shows the significant difference between deprived and prosperous groups.

Table 11.4 Relationships of glycaemic control among treatments with social deprivation categories

	Deprived	Intermediate	Prosperous	P value#
HbA1(%) in Insulin treated patients	11.19±1.97	10.13±2.19	8.85±3.89	0.0002
HbA1(%) in diet treated patients	9.34±1.99	9.28±1.25	8.7±2.89	0.434
HbA1(%) in tablet treated patients	10.08±2.98	9.83±2.6	9.59±3.3	0.205

#T-test and p value shows the significant difference between deprived and prosperous groups.

Table 11.5 Relationship of social habit with deprivation categories.

	Deprived*	Prosperous*	P value#
Smoking	114 (63)	26 (44)	0.02
No smoking	67	33	
Alcohol	109(60)	44(74)	0.04
No alcohol	72	15	
Total	181	59	

*Data is n(%)

χ^2 test and p value shows the significant difference between deprived and prosperous groups.

Table 11.6.a Relationships between complications of diabetes in 1995 and social deprivation categories.

Age adjusted prevalence of microvascular complications (1995)

Complications	Deprived (%)	Intermediate (%)	Prosperous (%)	P value[#]
Retinopathy	42.68	38.56	40.45	0.83
Neuropathy	51.48	13.47	20.02	<0.001
Proteinuria	57.02	25.60	21.68	<0.001
Ulcer/amputation	16.24	4.45	7.29	<0.001
Total	181	92	59	

#χ^2 test and p value shows the significant difference between deprived and prosperous groups.

Table 11.6.b Relationships between the number of complications of diabetes in 1995 and social deprivation categories.

Complications	Deprived (%)	Intermediate (%)	Prosperous (%)	P value[#]
complication - 1	67.23	31.15	29.62	<0.001
complications - 2	22.36	13.71	16.47	0.43
complications - 3	35.01	7.09	9.09	<0.005
Total	181	92	59	

#χ^2 test and p value shows the significant difference between deprived and prosperous groups.

Table 11.7 Relationships between age-sex adjusted mortality rate and social deprivation categories.

Age (yr.)	Deprived area (UPA>=30) Number of persons years followed	Number of deaths	Deaths/ 100/yr.	Prosperous area (UPA<=10) Number of persons years followed	Number of deaths	Deaths/ 100/yr.	Expected number of deaths among patients from prosperous areas if rates were the same as those of the patients from deprived areas
Males	n*=1898			n=173			
15-24	341	0	0	76	0	0	0
25-34	734	4	0.54	156	0	0	0.85
35-44	1249	7	0.56	262	1	0.38	1.47
45-54	2220	18	0.81	293	1	0.34	2.38
55-64	3623	73	2.01	384	7	1.82	7.74
65-74	2649	128	4.83	213	9	4.23	10.29
75-84	998	97	9.72	74	4	5.41	7.19
85+	108	16	14.81	8	1	12.50	1.19
Females	n=1758			n=122			
15-24	399	2	0.50	117	0	0	0.59
25-34	872	2	0.23	197	0	0	0.45
35-44	839	4	0.48	160	0	0	0.76
45-54	1777	13	0.73	180	0	0	1.32
55-64	2987	38	1.27	263	4	1.52	3.35
65-74	2566	86	3.35	196	7	3.57	6.57
75-84	1192	92	7.72	44	1	2.27	3.40
85+	205	25	12.20	9	0	0	1.10
Total	22759	605	2.66	1049	35	3.34	48.63
SMR			1.0			0.72 (35/48.63)	
Age-sex Adjusted Mortality Rate*			2.66			1.91 (0.72×2.66)	

* "n" is number of persons being followed up in the diabetic clinic at St Thomas' Hospital (1980-1994. Age-sex adjusted mortality rate is calculated using person-years method.

Table 11.8 Relationships between causes of death and social deprivation categories.

Underlying of cause of death (ICD-9)	Deprived*	Intermediate*	Prosperous*	P value#
Ischaemic Heart Disease (410-414)	356(9.74)	141(8.41)	11(3.72)	0.001
Cancer (140-208)	176(4.81)	59(3.52)	6(2.03)	0.01
Cerebrovascular disease (430-438)	99(2.71)	36(2.15)	6(2.03)	0.42
Circulation disease (390-459)	60(1.64)	24(1.43)	2(6.78)	0.40
End stage Renal Failure (584-586)	4(0.11)	0(0)	1(3.39)	0.11
Diabetes Coma (2501)	4(0.11)	0(0)	1(3.39)	0.15
Accident (E800-E999)	11(0.3)	5(0.29)	1(0.34)	0.99
Total@	3656	1676	295	

*Data is n (%);

χ^2 trend test and p value shows the significant difference between deprived and prosperous groups.

@ Data are total number of patients being followed in each area.

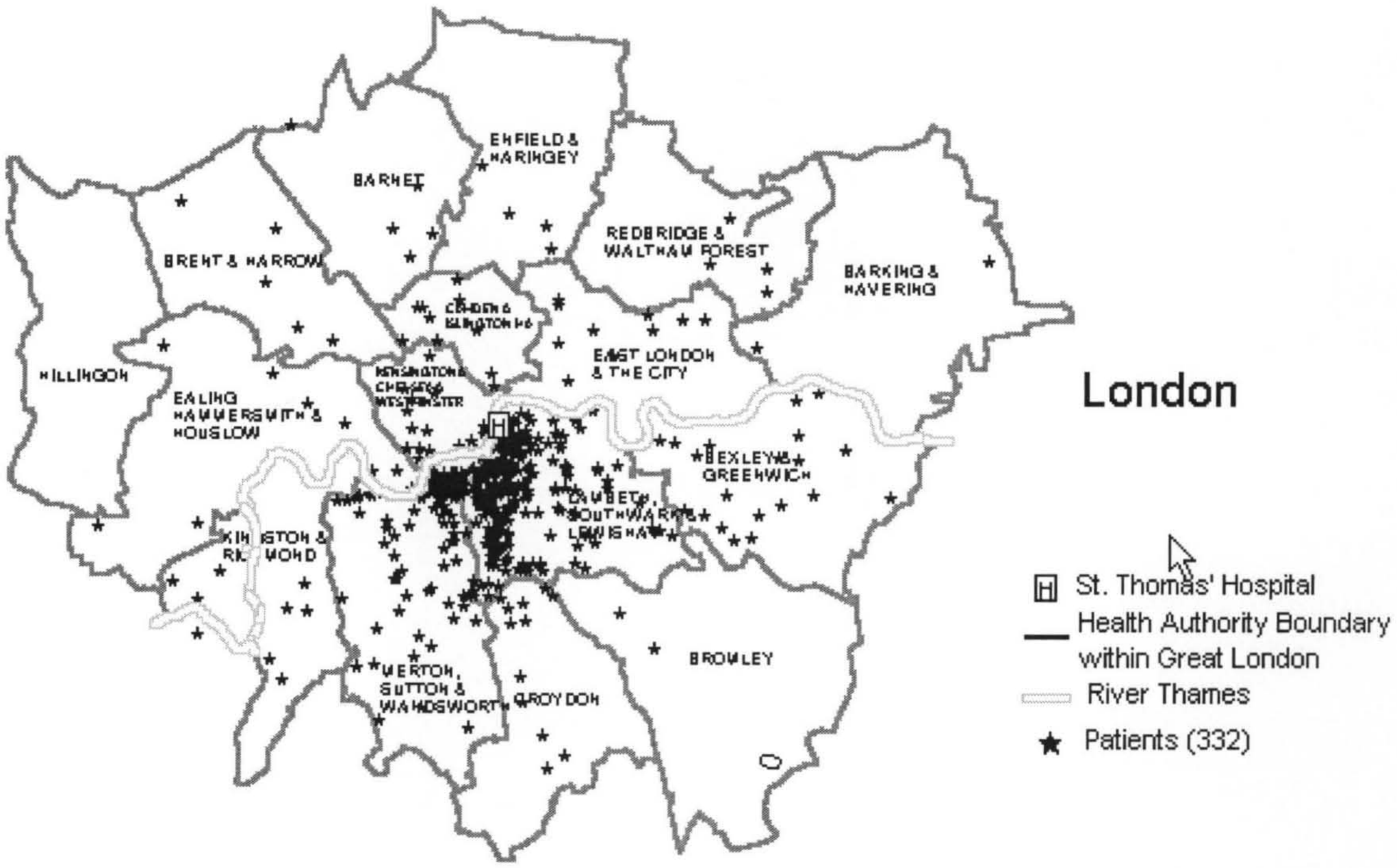


Figure 11.1 Geographic information about patients with diabetes mellitus attending the diabetic clinic at St Thomas' Hospital.

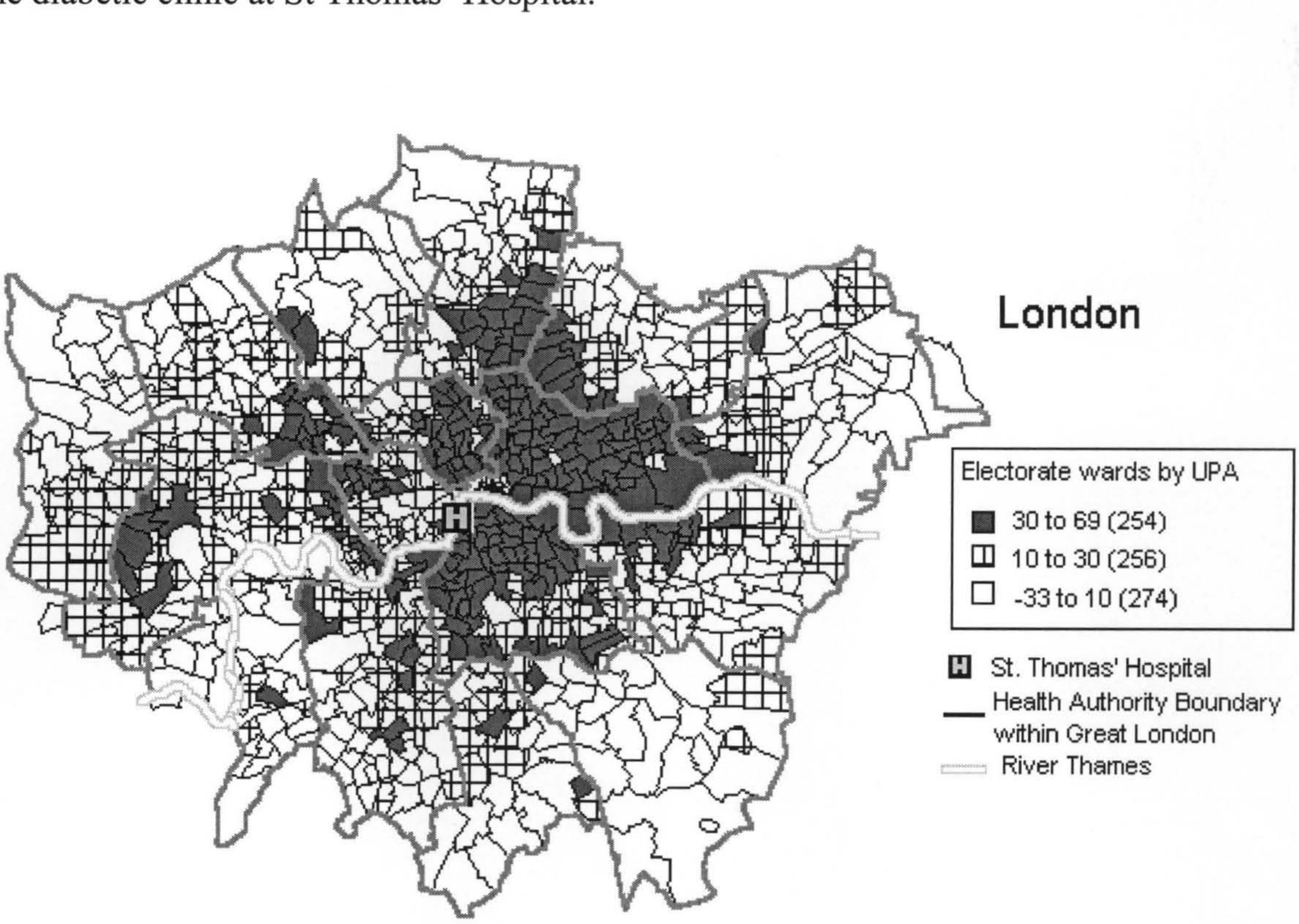


Figure 11.2 A pattern of social deprivation in Greater London.

Diabetic patients living in deprived inner city areas were generally older and were living alone. They had overall worse glycaemic control and were less likely to be on insulin. The use of insulin probably depends on several factors including the wishes and perceived compliance of patients. The minority ethnic groups were mainly Afro-Caribbean patients (78%) who lived in the deprived areas (e.g. Brixton in Lambeth area is known as one of the largest Afro-Caribbean communities in London). However the Caucasians still was the largest proportion population of in deprived areas compared to Afro-Caribbeans. This may indicate that Afro-Caribbeans might not be as bad on their diabetes control as Caucasians. As mentioned in the previous Chapter, prevalence of neuropathy was less in Afro-Caribbeans than Caucasians, and the age-adjusted mortality rate in Afro-Caribbean was lower than Caucasians (Table 10.3).

Similar results were repeated in Kelly's study on the Middlesborough-based population using the Townsend deprivation index (Kelly WF et al 1993/1994), in which a small proportion (5.3%) were from minorities, mainly from India and Pakistan. Connolly's study (Connolly VW et al 1996) showed that areas of low socio-economic status (social class) are at increased risk of cardiovascular disease in people with diabetes. Standardised mortality ratios in diabetic patients in the South East Thames Region are known to have a noticeable trend according to UPA score (Sönksen PH et al 1993). Although this study showed no difference in the prevalence of retinopathy between prosperous and deprived areas, Eachus et al report a strong link between social deprivation and diabetic eye disease (relative index of inequality 3.21; 95% confidence interval 1.84 to 5.59) (Eachus J et al 1996).

Patients living in deprived wards had more clinic visits, and worse diabetes control. This suggests that access to health care is clearly not a limiting factor, which is reassuring. On the other hand, the conventional approach on achieving good diabetes control in this group is ineffective. A controlled study carried out in this population also failed to show any beneficial effects of what on metabolic control, although being tailored specifically to their needs (Mulrow C et al 1987).

It seems likely that patients living in the deprived inner city find it more difficult to translate education on diabetes care into appropriate action and to comply with adequate advice on their diet. Another likely important factor is the 'lack of motivation' associated with loss of perceived 'locus of control', which may still stem from poor socio-economic status has been well documented in patients with poorer control (Shillitoe RW 1988). This study was based on referrals to a teaching hospital clinic. This may have produced a referral bias related to various factors, such as social class, unemployment, difficulties of access to diabetic services because of disability, poor education, and inadequate primary care schemes. The 'EURODIAB IDDM complication study' has clearly illustrated that the prevalence of severe microvascular complications is lower in educated men with Type 1 diabetes (Chaturvedi N et al 1996). There is less clear-cut information in Type 2 diabetic patients and the issue needs further investigation. A study on the contribution of nutrition to inequalities in health (Philip W et al 1997) has demonstrated that the diet in lower socio-economic groups provides cheap energy from foods such as meat products, full cream milk, fats, sugars, preserves, potatoes and cereals. The intake of vegetables, fruit, and whole-wheat bread is restricted in these people because of lower income and lack of appropriate 'health beliefs'. Our results suggest somewhat controversially, that contrary to established dogma, increased alcohol consumption may be positively related to good diabetes control and may negate the adverse effects of smoking (Table 11.5). If this is confirmed by further studies, then more rational advice on alcohol intake may be given to patients and might improve quality of their lives.

Unemployment, poverty and substandard housing have all been associated with a higher incidence of respiratory and cardiac disease, with a reduced life expectancy (Marmot MG et al 1986; Will JC et al 1988). We have shown that inner city deprivation is linked to increased morbidity and premature mortality in people with diabetes. Poor glycaemic control, with reluctance by patients to use insulin, and chronic hyperglycaemia even when using insulin, may contribute to the excess morbidity and premature mortality of diabetic subjects living in the inner city.

In conclusion, environmental factors affect diabetes outcomes; increased morbidity related to diabetes complications and an increased premature mortality rate in

diabetic patients is related to socio-economic and ethnic status. In such patients, obesity, poor glycaemic control and smoking habits are major risk factors for such events.

Although many different methods have been used to identify deprived populations of patients, (such as the Jarman Indices, Townsend Indices, Carstairs Indices and Social Class etc.) (Jarman B et al 1984; Townsend P et al 1986; Jarman B et al 1991; Main J et al 1991; Chase HD et al 1991), the adverse effect of social deprivation on certain outcomes of diabetes care (e.g. diabetes control and complications.) has been consistently demonstrated. The objectives of the St Vincent Declaration will not be achievable on patients living in deprived areas unless some new initiative is undertaken to tackle the problem in the new millennium. Clear, controlled and forceful planning, additional targeted resources and better organisation and management are required.

11.3 Conclusions

This pilot study (based on a cohort of diabetic patients) suggests that patients with increased morbidity and mortality are most concentrated in areas of lower social and economic welfare. This study could be easily done on the entire "Diabeta" database, if "Diabeta" provided an accurate and complete patients' current postcodes and clinical outcome measurements. Combining the GIS software (MapInfo) with "Diabeta" will allow us to utilise the data obtained from the link ("Diabeta"-NHSCR) more widely to explore the socio-economic and environmental factors related to diabetes and its complications.

Chapter 12

Evaluation of the feasibility of the record linkages between "Diabeta" and other health services information systems

The author has demonstrated that a record linkage between "Diabeta" and the NHSCR has resulted in valuable information on premature mortality and excess morbidity in our diabetic population. However this link still cannot enable measurement of the outcome of diabetes care in terms of use of NHS resources. This is due to the fact that clinical activity data on "Diabeta" is primarily about long-term out-patient care of diabetes at St Thomas' Hospital and contains only limited amounts of information on hospitalisation, emergency care, laboratory reports, home care etc. Therefore, a cost-effectiveness study based on "Diabeta" alone will be meaningless. The feasibility of further record linkages between "Diabeta" and other relevant information systems within the NHS has been discussed in Chapter 3 and showed that it is possible to link them with "Diabeta" in respect of technique, cost and confidentiality aspects. Further consideration, on which resource from these information systems is worthwhile/valuable to be linked with "Diabeta", will be tested in a pilot study, shown below.

In an attempt to identify the resources most likely to be consumed by our patients, a pilot study was carried out. A cohort of IDDM patients' identification data (n=126) was sent to all relevant information systems within the NHS (e.g. CMMD in Guy's & St Thomas' Hospital, CMDS in LSL, A+E own computer system in St Thomas' Hospital, "Proton" in Guy's & St Thomas' Hospital, RRS in STH, local GPs, "Comcare" in Lambeth Community Care Trust, Social Services system in Lewisham Council) for records-matching processes.

The number of successful matches from the cohort of records are more likely to represent the volume of each service consumed by the patient, rather than the accuracy of records matching. The accuracy of possible links was not studied in this

project, because of lack of time. The practical issues on "which resources should be linked to "Diabeta", and the possibility developing a cost model for diabetes care are discussed below.

12.1 Results of a pilot study on health care resources consumed by the cohort of diabetic patients

The characteristics of the cohort of diabetes patients (n=126) entering this study is described in Chapter 5.2 in "Methodology". The main aim of this study was to obtain information on the resources consumed by IDDM patients. The possibility of each resource used was measured by comparing the number of records detected[1] (after the linkage), with the maximum number of records expected to be detected by records matching procedure (Table 12.1).

Table 12.1 Result on matching cohort of patients' records between "Diabeta" and other relevant information systems.

Sources of the data	**Data to be linked into "Diabeta"**	**System Name (time period)**	**Number of records expected to be detected due to the linkage (a)**	**Number of records detected after the linkage (b)**	**The proportion of resource consumed by patients from the cohort (c) = (b)/(a)*100%** (95% confidence interval)	**Number of events detected after the linkage**
Secondary Care	Hospital in-patients episodes	CMMS (STH) (1990 - 1997)	126	33	26% (18%-34%)	104 episodes
		CMDS (LSL) (1987 -1997)	25	8	33% (15%-51%)	16 episodes
	A+E attendance in STH	A+E own computer system (1992 -1997)	126	26	21% (14%-29%)	50 attendance
	Number of dialysis in STH &GUY	"Proton"	126	0	0	0
	Laboratory tests in STH	RRS (Mar-Sep1997)	126	51	40% (31%-49%)	127 Lab. tests
Primary Care	Visits to GP & Prescription	GPs responses	72	43	60% (49%-71%)	
Community Care	Home visits by district nurses in Lambeth area	"Comcare" (1996-1997)	15	2	13% (4%-22%)	24 visits
	Disability (blindness) in Lewisham area	Social Services own computer system (1990 -1997)	1	0	0	0

[1] "detect" in the text means that number of records from the cohort of diabetes patients can be matched with the records from the corresponding NHS systems based on patient's identification data, and ascertain what services the cohort patients have consumed.

Additionally, in order to assess the proportion (p) of our diabetic patients that consumed each resource annually, the author also carried out the following calculation. Totalling the number of patients from the cohort who had an attendance for each service (e.g. hospitalisation, accident and emergency, primary care etc.) and each year (over the period of time that the record system was available to record the health care activities). Then using the mean value divide the total number of patients from the cohort who have attended at least once in the corresponding service over the period of time (obtaining this value from (b) in Table 12.1), and multiplying by 100. Confidence intervals (95%) were estimated ($p \pm 1.96 \sqrt{p(1-p)/n}$) for all proportional values.

12.1.1 Results of data matching on CMMS (STH & GUY)

The results on data matching has shown that 33 out of 126 patients with a hospital number extracted from "Diabeta" were successfully linked with records in CMMS, and 104 in-patient consultant episodes were provided on these matched records. This represented 26% of the total number of records in the cohort that had hospital in-patients episodes at STH and GUY's Hospital between 1990 and 1997. Therefore the possibility of admission to STH or GUY for IDDM patients for 7 years period, are at least 26%.

In this cohort, average of 6 patients/year were admitted (at least once) to GUY & STH. Approximately 4.8% (95% CI: 1.1%-8.5%) of our diabetic (6/126) patients had an in-patients admission in GUY & STH each year.

Seeing that the in-patient resource consumed by this cohort of IDDM patients is high, the author suggests that a record linkage between "Diabeta" and CMMS is not only feasible, but also necessary.

12.1.2 Results of data matching performed on Contract Minimum Data Set (CMDS) (LSL)

In the cohort (n=126), there were only 25 patients living in the LSL area. The 25 records were selected via the GIS software along with patients' local identifiers (date of birth, gender and postcode), and sent to LSL HA for data matching on their records in CMDS. This was to ascertain the total number of hospital in-patients episodes on patients living within our local health authority area.

The results of data matching on CMDS shows 8 out of 25 patients were successfully matched with records on CMDS. This represented 32% of patients who are living (lived) in the LSL area had at least one admission in a local hospital between 1987 and 1997.

The author estimated that approximately 8% (95% CI: 2.6%-18.7%) of our patients (2/25) living in the LSL area have one in-patients admission per year.

In conclusion, if these admissions include those not in STH & GUY, there will be an increase in the number of hospitalisations by the cohort of IDDM patients. For those patients living outside LSL HA area, a similar approach can be applied by sending the patients' local identifiers to the corresponding health authority. Ideally by sending identifications to the Department of Health (hospital episode section), all patients' hospital episodes wherever they occurred in this country may be detected.

12.1.3 Results of data matching performed by A+E computer system in St Thomas' Hospital

The results on data matching on the A+E system shows that 26 out of 126 patients in the cohort study were successfully matched. On those matched records, 50 re/visits in A+E were detected. 21% of the attendance occurred between 1992 - 1997.

It was estimated that 5.6% (95% CI: 1.2%-10.1%) of our patients (7/126) had episodes in A& E at St Thomas' Hospital every year.

Patients living within the St Thomas' Hospital catchment area are more likely to attend A+E at St Thomas' Hospital. Therefore the amount of A+E resource consumed by local IDDM patients may be as high as 32% (26/82) of patients living within the St Thomas' Hospital catchment area (<15km radius) between 1992 and 1997.

7.3% (95% CI: 4.4%-10.2%) A+E attendance (at least once) per year (6/82), owing to the fact that there were only 82 patients out of the cohort (n=126) living in the catchment area.

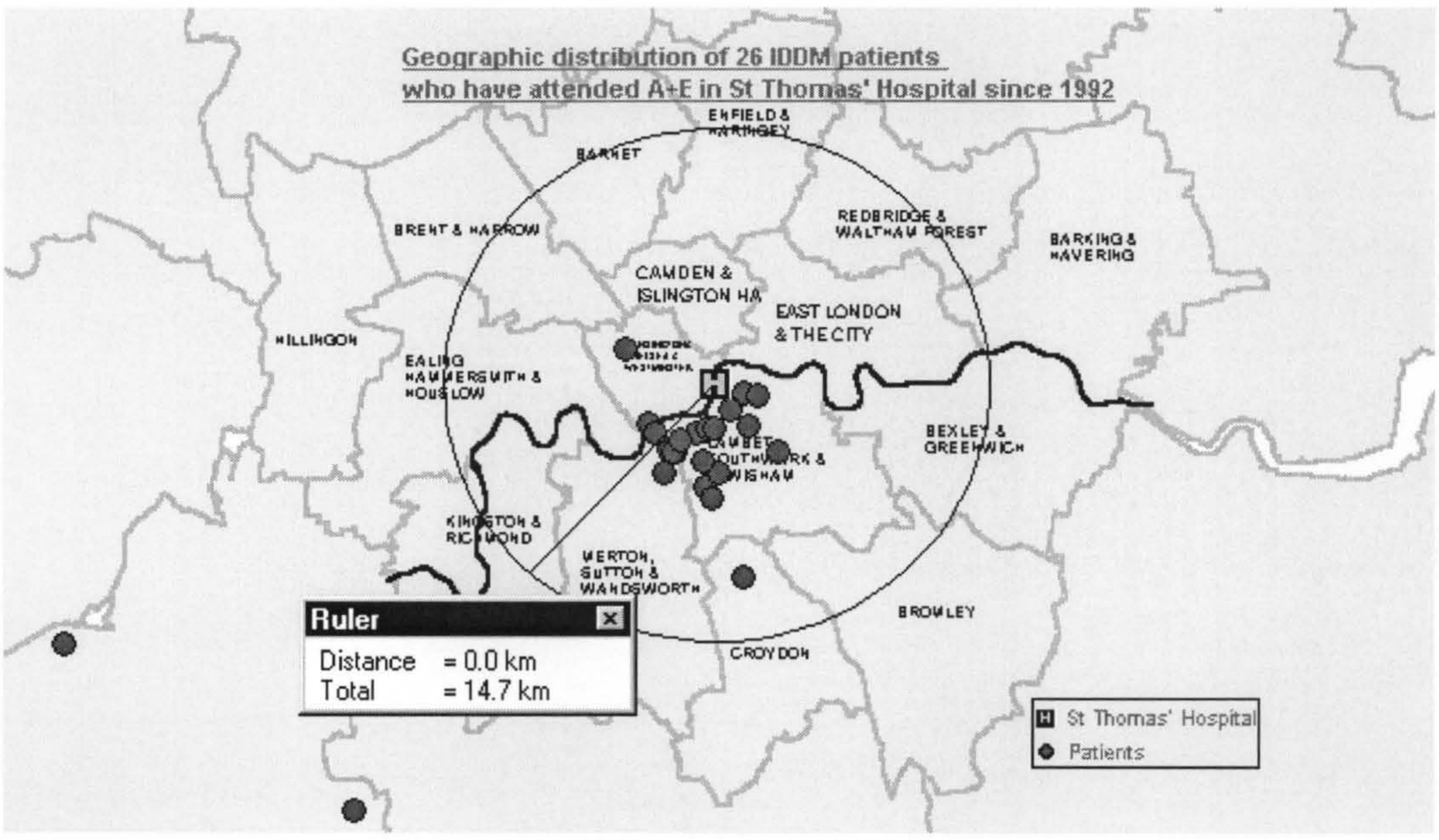


Figure 12.1 Geographical distribution of patients who attended A+E within the St Thomas' Hospital catchment area.

12.1.4 Results of data matching performed on "Proton" system in St Thomas' and Guy's Hospital

No sample records were matched, suggesting that no patients from the cohort were treated in the renal department for renal failure between 1994 and 1997. This most likely relates to the small size of our cohort.

12.1.5 Results of data matching performed on RRS in St Thomas' Hospital

Out of 126 records, 51 (40%) were linked (matched) with records on RRS, and 127 laboratory reports provided on the matched records. 40% of patients from the cohort had Lab. tests (chem, haem, micr, immu and viro) at STH & GUY for the previous 6 months (since March 1997). It was estimated that 8.7% (95% CI: 3.8%-13.6%) of our patients (11/126) have at least one laboratory test per month.

12.1.6 Results on letters to GPs

126 Letters asking the GP to provide information on patients' treatment, hospital/special (district nurse) referral and prescriptions were sent to each patient's GP. **43** GPs responded with either paper records or computer printouts. **13** GPs notified us that patients no longer attend their practices, and for those GPs (n=70) who didn't reply, the author contacted these GPs by telephone. **29** patients were still registered with the GP. 49 GPs could not be contacted within the time scale of this project. The overall response to our request from GPs (that patients are currently registered with) was 60% (43/72).

Those patients who had changed their GP since the last clinic visit or who have the incorrect address recorded on "Diabeta", could be traced through letters sent to the corresponding HAs as indicated by the NHSCR (experience from Dr Coppini's study shows that this is effective).

The information on GP systems indicates that most patients in the cohort regularly attend follow-up visits at GUY & STH. Out of 43 patients (whose records were supplied by GPs), 5 attendance out of 14 were in other accident and emergency department; 10 patients were referred to a medical specialist (e.g. local chiropodist, ophthalmologist etc.) and 7 patients to other hospitals. In conclusion, GP systems

can be used to identify the resources consumed by our IDDM patients outside STH & GUY, but these links are mainly manual at present and require significant resources to extract the data.

12.1.7 Results of data matching on the "Comcare" system in the Lambeth Community Care Trust

Matching shows that 2 out 15 patient records were linked (matched), and 22 re/visits by district nurses were detected, showing that 13% patients from the Lambeth area have had home care provided by a district nurse between 1996 and 1997. The volume of this resource in a year seems relatively high overall, compared to the utilisation of other resources.

12.1.8 Results of data matching on the Social Services system in Lewisham Council

The Social Services system in Lewisham Council was selected for the study because of their more advanced computerised records system and willingness to help. No records from the cohort could be matched on the Social Services System, which is no surprise since the number of patients from the Lewisham area in this study sample is very small (n=10). However the results from a previous pilot study on records linkage between the Social Services system in Lewisham and "Diabeta" shows that 33 out 200 patients records (16%) living in the Lewisham area could be matched with records in the Social Services system. Valuable resources from Social Services were not highly consumed by the cohort of patients studied.

12.2 Conclusions of results on the cohort study

People with diabetes form a significant proportion of patients who are admitted/re-admitted to a hospital or who re/attend A+E, and it is therefore important to link these systems (CMMS and A+E in STH & GUY, CMDS in the DoH) with "Diabeta" for an economic study on diabetes.

It is possible to collect more relevant patient information, particularly those related to services outside Guy's and St Thomas' Hospital from the patients' GPs, but this process is both time consuming and costly (especially for indirect/manual linkage).

"Comcare" and RRS computer systems are not suitable to be linked to "Diabeta" for an economic modelling study in a retrospective way, owing to the relatively recent introduction of computerised records or a limit time-period of computerised records. In a prospective study however, the amount of resources consumed by the cohort of IDDM patients indicates that these information systems should be linked into "Diabeta".

Data that cannot be directly linked into "Diabeta" includes mainly non-computerised records (historical "Comcare" and Social Services etc.). Indirect linkages are time consuming and will cost more.

Out-patient activities outside our diabetic clinic, such as visits related to vascular specialist or nephrologist, ophthalmologist in communities care cannot yet be determined using "Diabeta" via an "automated" linkage. These resources consumed by our diabetic patients are probably quite small (except ophthalmologist), and may be picked up via linkages with GP systems.

The possibility of developing a valid cost model based on all possible links will be discussed in next Chapter (13.3).

Section Four

Discussion and Recommendations Arising from the Project

The present study shows that an "electronic" record linkage between "Diabeta" and the NHSCR is not only possible, but is also desirable. Other linkages (direct/indirect) between "Diabeta" and other information systems are also feasible. Although potential benefits on all possible linkages have been demonstrated, the accuracy and efficiency of the linkage will be further discussed in this section. The validation of epidemiology studies based on the existing linkage ("Diabeta" - NHSCR) is also assessed. Finally, for the benefits of further development the author will identify possible weaknesses of the record linkages.

Chapter 13

Evaluation and validation of the record linkages

13.1 Errors existing in the linkage ("Diabeta" - NHSCR)

The NHSCR hold 64 million records on residents in England & Wales in its computerised registration system (by April 1999). To match records on "Diabeta" successfully and accurately with the massive database in the NHSCR, requires that the quality of data supplied to the NHSCR is very high. Insufficient discrimination power, noise, and missing data are major obstacles to the linkage of records (Acheson ED et al 1968). In results of "*linkage of records in "Diabeta" with the NHSCR database*" (Figure 8.2), 8.7% of patient records in "Diabeta" could not be matched with records in the NHSCR. This may be due to inadequate or inaccurate demographic data recorded on "Diabeta". Reasons for such inaccuracies include incomplete and inaccurate patient identification information in GP referral letters, especially due to unclear hand-writing (Appendix $C_{1\text{-}2}$), as well as the relatively common problems of patients sharing the same names, often seen in some ethnic groups. Even if patient identification details are correctly entered on "Diabeta", demographic data (name, address) cannot be updated accurately over time, unless the patient or GP notifies such changes. Little can be done to prevent such inaccuracies. Despite notices in the waiting area, patients often forget to notify the receptionist about a change of address. In order to minimise this problem, the demographic screen appears automatically when a patient's record is opened. It is still not infrequent that the patient sees this in the doctor's consultation room and comments that they haven't lived at that address for some time. In order to recover some records that cannot be linked with the NHSCR directly, preliminary efforts at linkage of such records with a local FHSA may be helpful.

The FHSA registration database holds patients' address and GP details, and has better discriminating power than the NHSCR (e.g. it holds information on patient's current and previous address). Matching local patients' data in a local FHSA database is much quicker and more accurate than that performed in the NHSCR.

The patient registrations database in the Lambeth, Southwark and Lewisham FHSA (LSL FHSA) covers most of the diabetic population in St Thomas' Hospital and may be linked to "Diabeta" (Figure 13.1). In one study, we selected a sample of 104 patients from "Diabeta" in order to evaluate the feasibility of implementing a linkage between "Diabeta" and the local FHSA. These records have been traced both in the NHSCR and in the LSL FHSA. Results showed that the local FHSA registration database enabled tracing of some patients who could not be traced by the NHSCR. In this sample, over 50% of records could not be fully matched in the NHSCR, but had been successfully matched in the LSL FHSA registration database. Patients' vital status and demographic data can be confidently updated in this way, so long as the patient remains resident in the LSL locality, if they have moved out of the area then the only source of linkage is through the NHSCR.

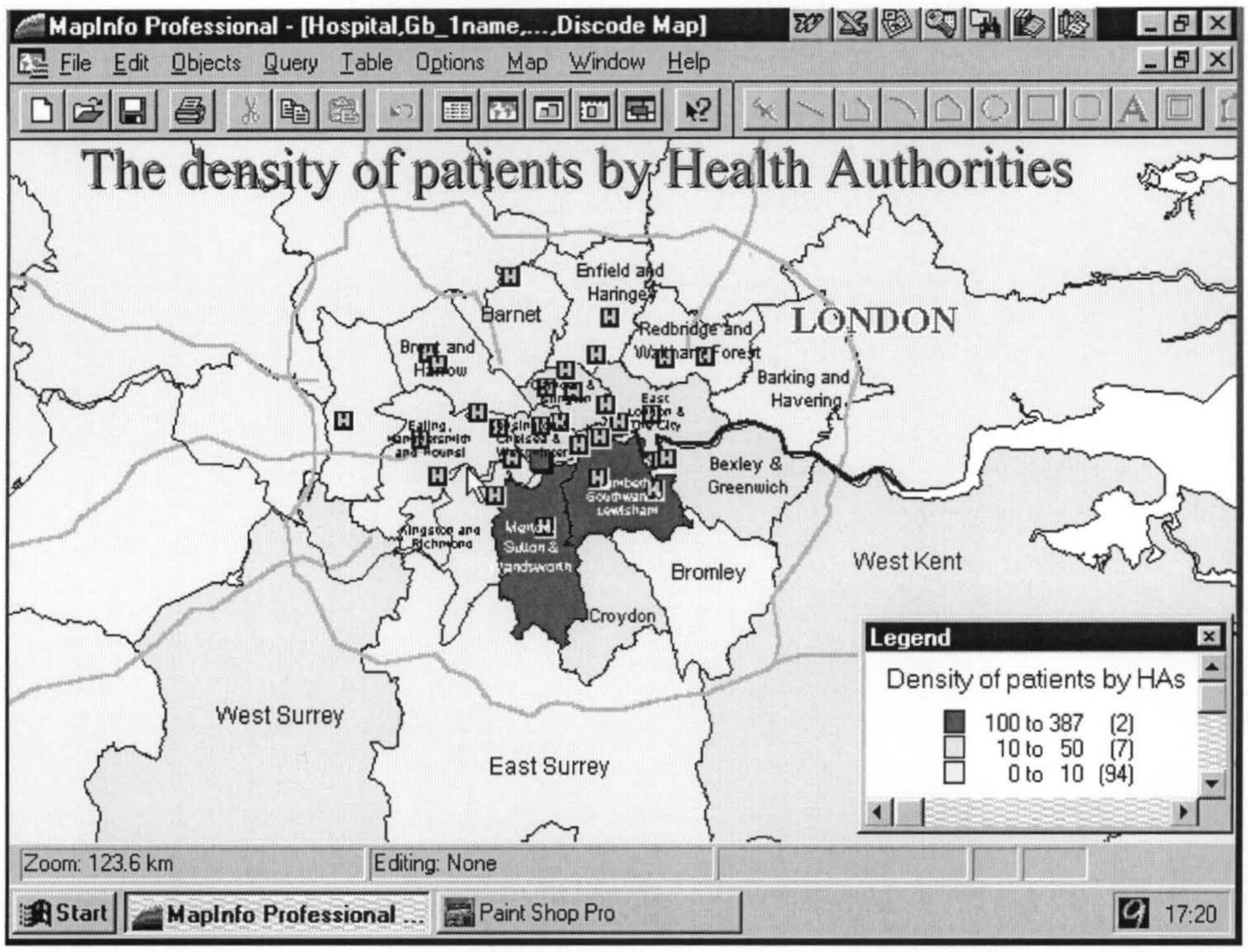
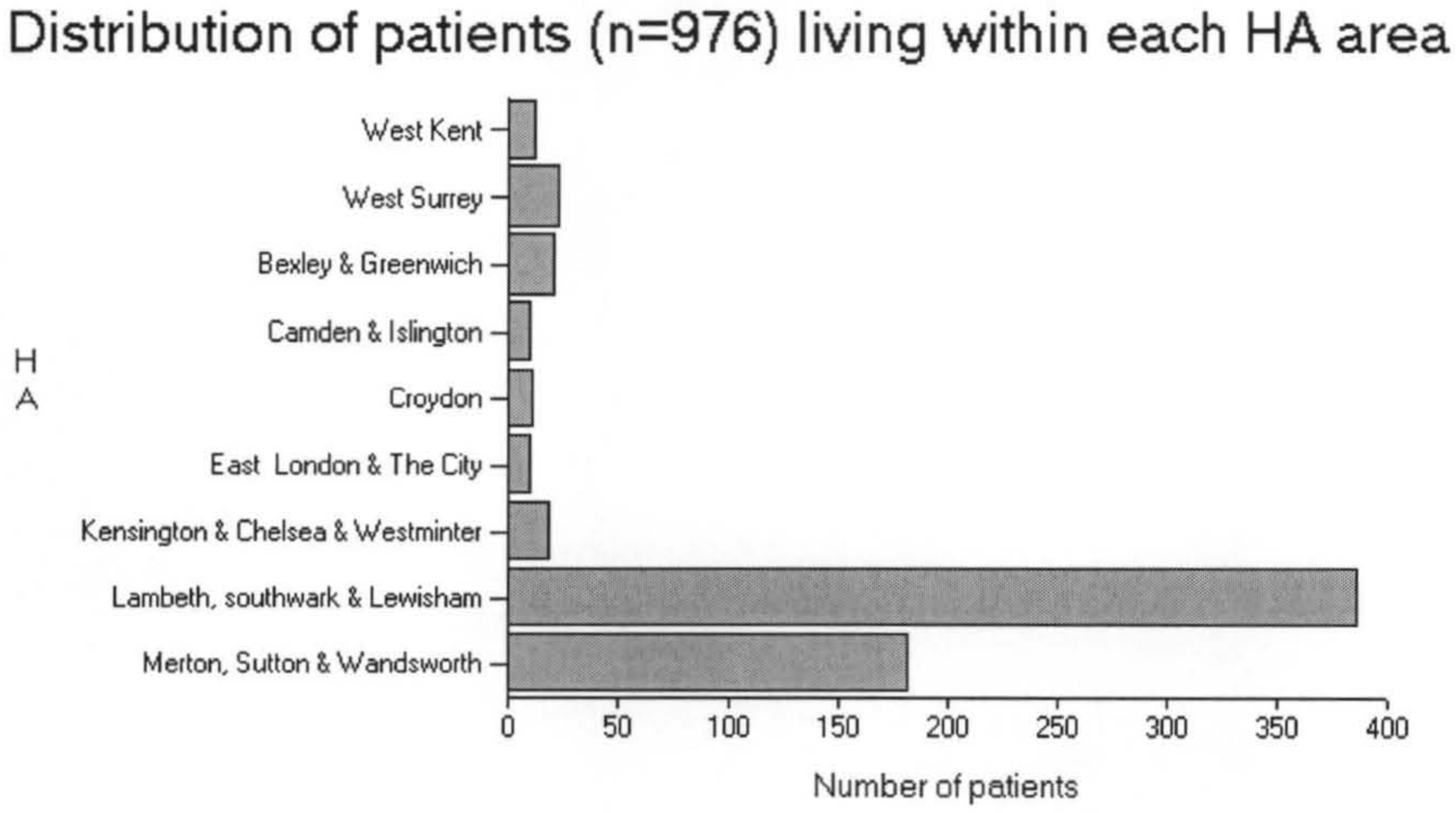


Figure 13.1 A distribution of diabetic patients within each FHSA.

(Map of density of patients by FHSA)



13.2 Validity of epidemiology studies based on the linkage ("Diabeta"-NHSCR)

Epidemiology studies on premature mortality based on the linked 'Diabeta' database have been illustrated by a representative diabetic population in the "Results" section (Chapter 9). The methods and results on mortality studies in diabetic population in various published studies are compared (Table 13.1).

Table 13.1 Mortality in diabetic patients compared to the general population in some published studies.

Author name*	Year	Geographic area	Age Group	Length of study	Size of cohort	Adjusted Risk Ratio							
						All-cause		Ischaemic Heart Disease		Cerebravascular Disease		Cancer	
						M	F	M	F	M	F	M	F
Sasaki et al [c,d③]	1988	Japan	all age	20	1,221	1.50	1.39	2.31		1.62			
Knuiman MW et al [c,d, ③]	1992	Australia	all age	8	888	1.83	1.43					0.79	1.09
Nyström L et al [b,d,②]	1992	Sweden	15-34	5	2,000	4.3	3.2						
Kleinman JC et. al. [a,d,③]	1988	Hyattsville	40-77	9	7,886	2.26	2.77	2.49	2.6				
Pan WH et al [a,e, ①]	1985	Chicago, USA	35-64	9	19,250			3.96	5.91				
Butler WJ et al [a,e,④]	1984	Michigan	>40	18	1,858			3.0	3.5				
Sprafka JM et al [b,d,①]	1993	Minneapolis, USA	>=45	10	540	1.26	1.44	1.38	1.99	2.33	1.46	0.65	0.94
Kannel WB et al [a,d, ①]	1979	Framhaming study, USA	45-74	20				1.7	3.3	2.2	2.2		
Walter DP et al [b,d,①]	1994	Southampton	all age	11	849	1.22	1.66						
Wong JSK et al [b,d,①]	1991	Aberdeen, UK	all age	5	4,186	1.04	1.2	1.41	1.61	1.01	1.14	0.65	0.82
Jarrett RJ et al [a,e,①]	1982	Bedford, UK	all age	10	249	1.10	2.76	1.42	5.16				
Fuller JH et al [b,d, ①]	1983	London, UK	all age	5	5,971	1.53	2.09	1.87	2.74	1.42	1.42	0.89	1.09
Present Study [b,d, ①]	present	London, UK	all age	15	7,542	1.14	1.47	1.66	2.04	1.29	1.38	0.78	1.05

* Four methods are used to determine the diabetic patients' vital status in mortality analyses:

① 'Flag'/'trace' records in the NHSCR or a similar population registration.

② Electronic linkage the local databases with a central register.

③ Using local computerised records system, questionnaire, self-report forms or medical records.

④ Using local newspaper and from individual questioning of local residents.

All the studies shown in Table 13.1 are comparison studies between either **diabetic and non-diabetic(a)** or **diabetic and general population(b)** or **NIDDM and general population(c).**

Risks of premature death in the diabetic population as all-cause of deaths or as specific causes of death, such as Ischaemic Hearth Disease (IHD), Cerebraovascular Disease (CVD) and Cancer (CAN) are assessed. The following measures are used in these studies:

(d) Standardised mortality ratio (SMR).
(e) Relative risk ratio (Cox proportion hazard ratio).

Eight mortality studies (62%) on diabetic populations used a population registration system (e.g. NHSCR). That was to ascertain patients' vital status (Fuller JH et al 1983, Jarrett RJ et al 1982, Kannel WB et al 1979, Pan WH et al 1985, Sprafka JM et al 1993, Wong JSK et al 1991, Walter DP et al 1994 and the present study). The studies in England (two studies in London and one in Southampton) have shown similar results to our study, except that the SMR for all-cause mortality in our female group was slightly lower than the other studies. All SMRs in the Aberdeen study were lower than in England. This may due to geographic differences.

In regard to other countries, only one study (Sweden) on their diabetic population used the electronic record linkage technique, and one (Minneapolis, USA) utilised a similar population registration system to identify patients' vital status. Results on mortality analysis from these studies were consistent with findings in our study. The SMR pattern for both males and females (in relation to cerebrovascular disease) in the Minneapolis' study was reversed as compared to the other studies, and the risk ratio in all-cause in Sweden's study was reversed also.

Other studies on premature mortality in the diabetic population were mainly based on local computerised records systems utilising a questionnaires/self-report/ physical medical records for confirming the patients' vital status. Most results from

these studies were also consistent with our findings, except that the Japan and Australia study showed that the SMR in NIDDM patients for all-cause mortality was higher in male than female group.

In conclusion, ascertainment of a patient's vital status has become an important issue for mortality studies in a diabetic population. The population registration system (NHSCR) has been recognised as a reliable system to determine a patient's vital status. However the electronic record linkage technique is still applied in only a small number of studies. The present study has shown that the SMRs were similar to the adjusted risk ratios (either SMR or risk ratio) in most published studies, no matter what methods have been used to identify the patients' vital status. In conclusion, the epidemiology study based on the linkage ("Diabeta"-NHSCR) is valid and useful. A record linkage behind a computerised clinical records system provides a valid and useful tool for epidemiology studies (particularly in mortality).

13.3 Validity of a cost-model based on all possible record linkages

In the relevant information systems, health care activity performance data, e.g. number of attendance or length of stay in hospital etc., are more complete than those of clinical data, e.g. diagnosis and operation procedures. Therefore it is important to identify initially where resources are used (such as episodes). The specific treatments related to the complications of diabetes (ulcer or amputation) can be relatively easy to find once the 'events' have been identified by using a coding system (e.g. Read Code) or through going back to the original records. The coding system for diagnosis and operation procedure is available in CMMS, CMDS, and some GP systems, but not in the Accident and Emergency system. The quality of coding obtained in CMMS and CMDS is reliable (Currie CJ et al 1997). Therefore, the cost of each service consumed by patients registered in "Diabeta" can be measured initially by the number of visits to general services (A&E, GP, Diabetic Clinic in St Thomas' Hospital etc.) and specialists. The estimation of cost on the complications of diabetes is possible once these 'events' are identified, e.g. tracing the records and investigating details of care activities on the complications in retrospective way. A more compete model is possible, but at present, both are

time consuming and expensive. We can conclude that the data sources are sufficiently rich to allow the construction of a valid cost model, although some costs will have to be estimated.

Because the payment of services in the NHS is mainly based on a package contractual relationship between health care providers and purchasers, it is impossible to identify the real cost of individual services. Various methods have been used to identify the cost of NHS contributions to diabetes care (Leese B et al 1992; Gray A et al 1995; Currie CJ et al 1997/1998; Gerard K et al 1989). The cost contributions may be estimated by an average unit cost or the hour, which may be the average hourly paid salary for health professionals. The hourly costs for health professionals in the DEDC at St Thomas' Hospital have been studied in a project of BSc in Medical Informatics (Hull D 1995). Alternatively it can be estimated by the national average data from the Department of Health (e.g. averaging the cost of length of stay in hospital per day as £177 (1992 price)). For example, General Practice visits and prescriptions cost £17.81. The cost of a below knee amputation is £5060 and a rehabilitation programme cost £2200 (Gray A et al 1995). In this way, the actual cost of care can be determined by the quantity of services multiplied by the unit price and added together.

The direct costs of illness (such as drugs, therapy, consultations and hospital care) are more easily determined than indirect costs, which include the value of production lost to society due to illness (e.g. loss of earnings, early retirement and premature death etc.). Studies in the USA estimated indirect costs to be nearly half of the total cost (Warner DC et al 1996). In "Diabeta", after the vital status (alive/death) is ascertained through a record linkage ("Diabeta"- NHSCR), the total annual productive loss from premature deaths attributable to our diabetes population can be mainly determined through the remaining working years together with the average annual earning. The unemployment status for our patients can be determined by the field of 'occupation' in "Diabeta", however the quality of this information is dependent on reports from patients during their consultation and vigilance of the doctor in clinic checking whether it is up to date. In order to identify the unemployment status for an entire database, a linkage between

"Diabeta" and the Department of Social Security system (DSSs) based on a National Insurance Number may be a solution (Gray A et al 1995). The total annual productive loss from unemployment can be calculated in a similar way as for premature death, and the cost arising from unemployment benefits (e.g. income allowance and housing benefits) will need to be considered as well.

It is impossible to give an exact cost of all services consumed by diabetic patients. Assessing the cost of diabetes services should represent the main cost (e.g. those resources are most likely utilised by people with diabetes) to the NHS, and this may help the NHS allocate the necessary resources. In this study, the author has identified where the services are most needed (Chapter 12) in order to prevent high costs.

13.4 Weaknesses and strengths of record linkages

Information systems in hospital and community care are currently inadequate to monitor attempts to reach the targets of WHO St. Vincent Declaration (Sönksen PH et al 1993). This study has introduced a method to improve the accuracy and completeness of patient data recorded in a hospital information system through a record linkage technique. It enables an information system to identify the true 'final outcome' (clinical/socio-economic) of the diabetic population and to analyse risk factors related to excess mortality and morbidity in diabetic patients (Weng C et al 1996). The evaluation of diabetes care based on information systems in hospitals or community care settings can be misleading through inaccuracies in patients' vital status and incomplete clinical activities data recorded on them.

Implementation of this linkage ("Diabeta"- NHSCR) during this pilot study, has been very successful and provides a stronger foundation for maintaining continuous and high quality patient care. The further linkages between "Diabeta" and other relevant information systems within the NHS need to be developed for better monitoring of cost-effectiveness of services provided to the diabetes population.

Despite the advantages of the record linkage, patient data confidentiality has always been an important issue. Many population-based databases (e.g. NHSCR, FHSA

etc.) do not record clinical data. Along with the NHS reforms (e.g. internal market), information technology has become more important for monitoring the quality of services provided to patients, rather than for 'administrative' alone. Introduction of the NHS wide network and the replacement of the old NHS number with the new NHS number (a 10 random digital unambiguous number which ensures that it has no link with personal data), has made the exchange of clinical information within the NHS possible in terms of data security. The Nation Wide Clearing System (NWCS) database is one of a number of products developed since the NHS reforms in April 1998, and is a population-based database containing all hospital episodes in England and Wales.

Moreover, information from a death certificate is far more sensitive than patients' clinical data, mainly because of legacy issues. There are two computerised systems that contain information from death certificates in the Office for National Statistic (ONS). One is used for statistical purposes (excludes patient identifiers) and one is used for management purposes. The latter contains complete data from death certificates, but is not authorised to have an electronic linkage with any population-based database. For these reasons, the paper format death certificates will continue to be used for some time.

In Scotland and Wales, a Community Care Number (a unique patient's identification number) is issued to every resident and this is used in all health care systems, allowing various epidemiology studies are to be conducted on a large population basis. The recent epidemiology study in diabetes from Wales (Currie CJ et al 1996/1997) successfully measured the clinical and socio-economic outcome of diabetes care provided in the community.

As mentioned earlier, central London is a complex inner city with a large multi-ethnic community that is very mobile and relatively poor. These characteristics make delivering and monitoring of care provided to the diabetic population more difficult. However if the new NHS number is used in all health care settings, it will be possible exchange patient information (demographic & clinical) correctly and

safely between all health care settings and to obtain a more complete health care information on each individual patient.

Chapter 14

Recommendations arising from the project

14 .1 Considerations of how to achieve a complete update of patients data in "Diabeta" in the future.

Modernising Britain has been a central theme of the Government's programme since it came to office in May 1997. Crucial to that objective has been the drive to modernise the NHS, with the aim *of "giving the people of this country the best system of health care in the world*". Better care for patients, and improved health for everyone depends on the availability of good clinical information and accessibility when and where it is needed. The national information strategy proposed that the information needs to be person-based, and integrated with Electronic Patients Records (EPRs) in primary, secondary and communities care. The quality of data recorded in EPRs essentially needs to maintain **accurat**e, **complete**, **relevant**, **up-to-date** and **accessible** information. Such information has to be secure and confidential whilst it is shared across the NHS. ('*Information for health'* - An Information Strategy for the Modern NHS 1998-2005).

With regards to this study, the update of the patient's current address and GP details in the "Diabeta" database via a record linkage with the local FHSA database is limited to those patients registered in the local database. Meanwhile, the collecting of diabetes care activity data related to primary care (such as that provided by a GP, district nurse or social services staff via a local system) is also limited to those persons who are registered in the corresponding local systems. Although the information on drug dispensing of patients can be transferred from GPs' sources into "Diabeta", this is normally a tedious process. The question arises as to how the entire database can be updated and completed (maintaining lifelong electronic patients records) more efficiently and effectively in terms of patient demographics and clinical activity (including drug dispensing) data. This can hopefully be answered in the near future. The option of future development on "Diabeta" will be discussed below.

14.1.1 Linking "Diabeta" with the "DIALOG" programme.

Extensive research is currently on-going in the Family Health Services Computer Unit (FHS CU) regarding the feasibility of installing "DIALOG" in each FHSA (Vaughan NJA et al 1996). The "DIALOG" system is designed to interface with the FHSA registration database, enabling real-time downloading of demographic data (Vaughan NJA et al 1996). The intention is to register all diabetic patients (with patients consent) within a district. Its main aim is to support annual check-ups. Based on this selection, a reminder could then be generated and sent to the lead clinician (consultant or GP). A standard clinical data set (Vaughan NJA et al 1995) could be collected from the lead clinician on a ASCII file diskette, and then transferred onto "DIALOG".

A national diabetes register could be set up in the FHS CU. This register could hold registration details on all diabetic patients in the NHS by linking the "DIALOG" system run in each FHSA (Diabetes Care Management 1994). Once this is available, any change in patient demographic information is reported to the FHSA and will be continually updated on the diabetes register. If patients move out of the FHSA area, this would be notified to the new FHSA and the patient's latest annual review report will be directed to the new FHSA.

A future link between "Diabeta" and "DIALOG" could eventually lead to a complete update of patient registration and clinical details in the "Diabeta" database in a cost-effective way. The cost for updating any diabetic clinical information systems is, however, quite substantial at about £3,000 per annum (Vaughan NJA et al 1996) (Appendix D), but will help ensure that every diabetic patient has the opportunity to be reviewed on a regular basis. This linkage may make a substantial contribution towards reducing morbidity and mortality in our diabetic population (Diabetes Care Management 1994). Research in this area is incomplete, but pilot studies on the "DIALOG" system have already been performed in four FHSAs (Diabetes Care Management et al 1994).

One further point arises: once "DIALOG" is introduced at a general level, will the link between "Diabeta" and the NHSCR still be useful? The answer is probably yes, as only the NHSCR can provide details on the cause of death of patients. The "DIALOG" system will, on the other hand, only provide the date of death and its forecast costs are substantially higher. Further developments on "DIALOG" in order to provide copies of death certificates are clearly indicated.

Although the "DIALOG" system can provide updated demographic data, clinical examination details and partial data of diabetes care activity for patients registered in the system, it doesn't include all diabetes care activity data (e.g. all admissions occurring in patients, attendance to A+E or a specialist etc.). In order to solve this problem, further development on "Diabeta" is required.

14.1.2 Linking "Diabeta" with the NWCS database

As the NWCS database has only been established since April 1997, it may not be of much help for a retrospective cost-modelling study, but will be important for a prospective study. In theory, "Diabeta" can supply the new registers' records (which hold patients' local identifiers (new NHS number, Name, Date of Birth, etc.) to the NWCS database through NHS wide-networking. A 'flag' can be entered into the patients' records on the NWCS database to indicate that the records belong to "Diabeta". The NWCS database can inform "Diabeta" whenever a hospital episode occurred on patients' records and then send details of the consultant episodes to "Diabeta". If "Diabeta" is linked with the NWCS database, all hospitalisations occurring in England & Wales on 'flagged' patients could be notified to "Diabeta". As a result, the linkage between "Diabeta" and CMMS will only be necessary to obtain details on hospital episodes (e.g. specialist code etc.).

The contract data sets in A+E and community care are still in the range of a pilot study in a local area. In the future, contract data sets related to all health care providers (hospital, A+E and community care etc.), can be given to the NWCS database. Therefore "Diabeta" will be able to link most resources related to treatments provided for all our patients within the NHS.

14.1.3 Linking the Prescription Pricing Authority system

The Prescription Pricing Authority (PPA) collects prescriptions from all pharmacies and chemists, prescribed by GPs in England, and its function is to monitor drug dispensing within the NHS. This system only records a GP code as a patient's identification data and therefore cannot cross- reference prescriptions with individual patients. A special code (a 'flag') on GP prescription helps to identify if a prescription belongs to a particular group of patients (e.g. IDDM patients from "Diabeta" in St Thomas' Hospital). A monthly report on the cost of drugs used for our patients can be provided to us by the PPA researchers. The report can also produce the cost of dispensing prescriptions in relation to diagnosis, according to BMF (British Medical Formulation) code. The charge of a 'flagging' study will be 10 pence per record, per year.

In conclusion, instead of collecting prescriptions from each individual GP, the cost of drugs taken by a cohort of patients can be received from the PPA through a 'flagging' study by asking their GPs to enter a "flag" on the prescription for our patients. This may be useful for a prospective cost-modelling study, if the criterion of patients can be initially defined (e.g. IDDM or NIDDM).

Since there is increased demand to measure the outcome of drug treatment in the NHS, rather than monitoring dispensing, the PPA intends to assign the New NHS number to all who newly register with the PPA database in the future. If "Diabeta" is linked with the PPA records system, based on the patient's new NHS number, the cost of drug dispensing for patients registered on "Diabeta" will be easily monitored and measured with cross-references to patients' clinical outcome data.

14.1.4 Possibilities of introducing the existing links ("Diabeta"-NHSCR) to the NHS wide network.

As described earlier, existing communication across the NHS is only on the level of a regional network (e.g. LCS), district & supply network (Racal net) and an ad-hoc use of a dial-up link. The NHS needs to rationalise these supplies and develop an infrastructure to interlink the whole of the NHS in England. The aim of developing the NHS wide network (NHSnet) is ultimately to improve the health of the

population and provide better patient care (A handbook for IM&T specialists, 1996). The NHS-wide networking programme will achieve these links by creating a national networking infrastructure connecting all NHS organisations enabling exchange of both unstructured and structured electronic data/mail between various NHS settings. Major benefits include:

- It allows GPs to communicate with Health Authorities.
- It provides all NHS organisations with connections to national bodies (e.g. NHSCR etc.), including NHS suppliers.
- It enables information exchange between provider and purchaser.

The potential of the NHSnet to take the NHS forward into the new era of information technology has been recognised by the Government. The development of the NHS "Information Superhighway" is seen as a major advancement in the *New NHS* (*Information for Health*, An Information Strategy for the Modern NHS 1998-2005, NHS Executive). The NHSnet has the potential to deliver the enormous benefits of information technology to all parts of the NHS for a wide variety of purposes.

The NHSnet will be the best medium for the transfer of clinical information, however the power of the Internet in the global development of information technology is a particularly rich source of academic information. The NHS plans to take maximum advantage of the features of a managed service such as the NHSnet, and also accommodating developments within the wider Internet. Specifically, NHSnet already provides Simple Message Transfer Profile (SMTP) relay services through a safe gateway ("firewall") to the Internet and should be enhanced to provide an SMTP mail service. This provides the clinician with an Internet mail address as well as an X400 email identifier. As technology allows, these should be integrated to appear as one to the recipient. Figure 14.1 demonstrates how a clinical information system (e.g. "Diabeta") connected with the NHSnet, and how mails can be through Internet via NHSnet.

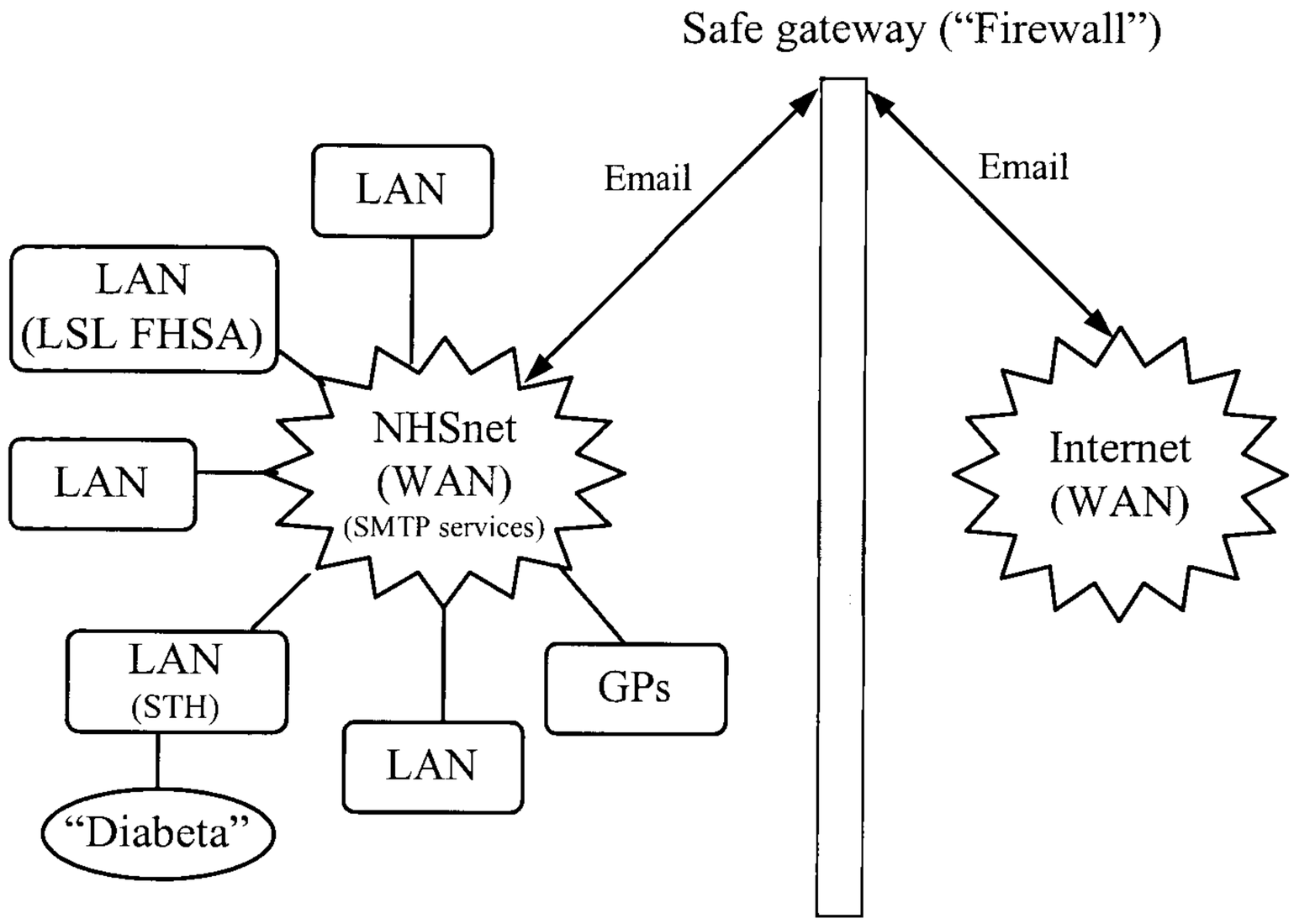


Figure 14.1 A diagram of the NHSnet showing its connection with Local Area Networks (LANs) within NHS organisations and interaction with the Internet.

In respect to security of patient's data, the Access Agreement and Code of Connection to NHSnet are continually reviewed by the IM&T, to ensure the balance between arrangements for security and accessibility is appropriate.

The cost of the NHS-wide network can be cost-effective, since the NHS-wide networking programme has negotiated new tariffs with telecommunication providers ensuring NHS organisations obtain discounts that reflect the size of the total NHS spend in this area.

In addition, Electronic Data Interchange (EDI) is a technology which has been shown to reduce errors, simplify procedures and improve the quality of computer data exchange. The EDI standard UN/EDIFACT (United Nations/Electronic Data Interchange for Administration Commerce and Transport) has been used to promote the use of EDI across all sectors of the NHS and can be break down as follows:

1) General EDI infrastructure and skills
2) National standard messages endorsed by the appropriate user community, e.g. clinicians, general managers.

As mentioned earlier, the EDI message (NHSCDS version 4 formats) used in NHSnet service has enabled providers to convert their local format to NHSCDS (e.g. CMDS) via National Wide Clearing Services (NWCS).

EDI requireds expensive software and special network connections, but with the advent of XML (eXtensible Markup Language) and the second generation Web, its usefulness to NHS is being re-examined. XML is a way of tagging any kind of data to make its significance an understandable event to machines. The tags created with XML resemble the HTML (Hypertext Markup Language) used today to create Web pages. However HTML tags generally indicate only how the content should appear, XML tags separate content from presentation, and indicate what the contents means.

If the medical profession uses XML to hammer out a markup language for encoding medical records, the records can be sent by e-mail which could contain <patient><name>blah blah </name><drug-allergy></patients>. Programming any computer to recognise this standard medical notation and to add these records to its database becomes straightforward (XML and the Second-Generation Web, 1999).

Just as HTML created a way for every computer user to read Internet documents, XML makes it possible to exchange information between incompatible computer systems. The application for XML in the years to come will be computer-to-computer communications instead of computer-to-human or human-to-human communications (http://www.research.ibm.com/ resources/ magazine/ 1999/ number_1/ xml199.htm).

In conclusion, the New NHS number, NHSnet and modern data exchange language should allow information exchange among all NHS organisations more efficiently,

cost-effectively, securely and will improve the quality of patient data in a local database.

14.1.5 Linking "Diabeta" with Supermarket Consumer Resources

According to the results shown in Chapter 11 ("*The identification of social and geographical factors related to increased morbidity and mortality in diabetic patients*"), the outcome of diabetic care is strongly related to people from low socio-economic backgrounds. One of the reasons could be the poor nutrition obtained by people from socio-economic deprived areas. In order to identify the association between nutrition, food, socio-economic background and the outcome of diabetic care in people with diabetes, linking "Diabeta" and supermarket consumer resources could be a solution. This is to see what kinds of food (and price of the food) are consumed in a geographic area for people with diabetes and if it is related to socio-economic factors. Most supermarkets (Sainsbury, Safeway and Tesco etc.) are encouraging their customers to apply for Loyalty Cards (or Club Cards) in order to obtain bonuses for purchasing their products. Persons' name, address (postcode) and products they purchased are recorded on their database. If "Diabeta" can be linked with such a data source, evidence on what, and how, food products are related to better or worse outcome of diabetes may be provided. On the other hand, how food products related to ethnic minority groups, in people with diabetes, may also be clarified.

14.2 The value of GIS technology in the planning and delivery of focused health care to diabetic subjects

The author has demonstrated in this thesis that an updated and complete hospital diabetes database can be achieved through a record linkage technique. The author also described how such a database has been used to identify the risk factors (clinical, social and economic) related to the outcome of diabetes care. There is however an urgent need to know, why and how, such risk factors can be prevented within the NHS budget. The "Diabeta" development needs to be re-targeted, so as to utilise existing resources more effectively in order to achieve a better outcome of diabetes care.

GIS technology can bring various data sources (e.g. patients' demographic/clinic data, census data, geographic data and diabetes facilities/resources data etc.) together in a geographic format. Cross-sectional analysis of these data will define the geographic areas of high risk of adverse outcome from diabetes. It can also identify and define inequalities of access to relevant services and determine other geographical and socio-economic factors influencing these. Further investigation on the benefits of linking GIS with a patient clinical database is recommended. The aim is to provide and evaluate a planning model to facilitate provision of adequate and appropriate resources for delivery of diabetes care for those most in need, in a given community.

If the study is successful, this project could introduce the use of GIS technology into health care planning and information distribution for diabetes.

In order to visualise where the diabetes service resources are located and what structured preventative care is provided for patients with diabetes in a local community, relevant information relating to services in primary, secondary and community care should be collected. This will include the number of community and hospital-based consultants, specialist medical registrars, medical research fellows, specialist diabetes nurses, other nurses, chiropodists, dieticians, eye-screeners, ophthalmologists, psychologists, psychotherapists and GP specialist clinics together with their staffing from the appropriate organisations. The number of sessions that each devote to diabetes will be recorded. The data will include the types of services for people with diabetes (e.g. education, supervision, support, estimation of weight, plasma glucose control, BP, Lipid management, eye and foot screening etc.:), quality measures used, interpretation facilities provided, frequency of clinics, attendance patterns, waiting times etc. Data from these facilities/resources can be overlapped with patient's address geographically via GIS. It should be relatively straightforward to develop a computer program that would allow nursing and other medical staff to retrieve useful data and help others to obtain adequate and relevant information. Examples might be: 'Where are the nearest places for the patient to receive treatment?' and 'What services may they expect, in a timely fashion?' This should also help the BDA and Health Authorities

evaluate whether diabetes clinics in various geographical areas meet their required quality standards. It will be necessary to provide this information in conventional publications and as part of a 'Diabetes and Endocrine' web page at Guy's and St Thomas' Hospital Trust so that it is on the both the NHSnet and the Internet. This will allow the growing number of diabetic patients with PCs at home to access such information.

References:

A handbook for IM&T specialists. Implementing the infrastructure for Information Management & Technology in the NHS. 1996.

Acheson ED. Record Linkage in Medicine. *E.& S. Livingstone Ltd. Edinburgh and London* 1968, 197-198.

Bair 1992. 'Running out of Steam', *Health Services Journal*, 11 June 1992, 26-7.

Baldwin JA, Achesoh ED and GRAHAM WJ. Textbook of Medical Record Linkage. *Oxford University Press* 1987, 6.

Balkau B and Papoz L. Certification of cause of death in French diabetic patients. *Journal of Epidemiology and Community Health* 1992, 46: 63-65.

Black D, Morris J N, Smith C and Townsend P. Report of the Working Group on inequalities in health. London. *Department of Health and Social Security* (DHSS) 1980.

Bloom S, Till S, Sönksen PH. Smith S. Use of a Biothesiometer to measure individual vibration thresholds and their variation in 519 non-diabetic subjects. *BMJ* 1984, 288: 1793-1795.

Borch-Johnsen K, Andersen PK, Deckert T. The effect of proteinuria on relative mortality in type 1 (insulin dependent) diabetes mellitus. *Diabetologia* 1985, 28: 590-596.

British Diabetic Association. St. Vincent Joint Task Force for Diabetes, *Department of Health* 1995.

Butler WJ, Ostrander LD, JR., Carman WJ and Lamphiear DE. Mortality from coronary heart disease in the Tecumsem study. Long-term effect of diabetes mellitus,

glucose tolerance and other risk factors. *American Journal of Epidemiology* 1984, 121: 4, 541-547.

Caddick SL, McKinnon M, Payne N, Ward TJ, Thornton-Jones H, Kells J and Ward JD. Hospital admissions and social deprivation of patients with diabetes mellitus. *Diabetic Medicine* 1994, 11: 981-983.

Carstairs V and Morris R. Deprivation and mortality: an alternative to social class? *Community Med* 1989, 11: 210 -219.

Chase HD and Davies PRT. Calculation of the underprivileged area score for a practices in inner London. *British Journal of General Practice* 1991, 41: 63-66.

Chaturvedi N, Fuller JH and Stephenson JM. The relationship between socio-economic status and diabetes control and complications in the EURODIAB IDDM complications study. *Diabetes Care* 1996, 19: 423-430.

Collett D. Modelling survival data in medical research. *Chapman and Hall, London* 1994.

Connolly VW and Kesson C M. Socio-economic status and clustering of cardiovascular disease risk factors in diabetic patients. *Diabetes Care* 1996, 19: 419-422.

Coppini DV, Weng C, Jones MC, Young PJ, Macleod AF and Sönksen PH. Natural history of peripheral neuropathy in relationship to vibration perception thresholds in diabetic patients. *Diabetologia* 1997, 1[Supp11] P2187.

Coppini DV, Weng C, Meeking DR, Abraham S, Young PJ, Sonksen PH. Log-transformation of VPT is useful in identifying the presence of neuropathy in diabetic subjects. *Diabetic Medicine* 1996, 13: [Suppl 7] P78 (Abstract).

Coppini DV, Young PJ, Weng C and Sönksen PH. Risk factors associated with the development of end stage peripheral neuropathy: a 10-13 year follow-up study. *Diabetic Medicine* 1996, 13 [Suppl 3] P85 (Abstract).

Coppini DV, Young PJ, Weng C, Macleod AF and Sönksen PH. Outcome on diabetic foot complications in relation to clinical examination and quantitative sensory testing: a case-control study. *Diabetic Medicine* 1998, 15: 765-771.

Currie CJ, Kraus D, Morgan CLl, Gill L, Stott NCH and Peters JR. NHS acute sector expenditure for diabetes: the present, future, and excess in-patient cost of care. *Diabetic Medicine* 1997, 14: 686-692.

Currie CJ, Morgan CLl and Peters JR. The epidemiology and cost of inpatient care for peripheral vascular disease, infection, neuropathy, and ulceration in diabetes. *Diabetes Care* 1998, 21(1): 42-48.

Currie CJ, Williams DRR and Peters JR. Patterns of in and out-patient activity for diabetes: a district survey. *Diabetic Medicine* 1996, 13: 273-280.

Data Protection Act. 1984. *London: HMSO.*

Department of Health 1989c. Term of Services for Doctors in General Practice. *London, Department of Health* 1989.

Department of Health 1991. Research for Health. A Research and Development Strategy for the NHS. *London. Department of Health.*

Diabetes Care Management. *First Health.* East Sussex FHSA, Summer 1994, 6-7.

Diabetes care and research in Europe. The St Vincent Declaration. *Diabetic Medicine* 1990, 7:360.

Diabetes Integrated Care Evaluation Team. Integrated care for diabetes: clinical, psychosocial, and economic evaluation. *BMJ* 1994, 308: 1208-1212.

Dorman JS, Laporte RE, Kuller LH, Cruickshanks KJ, Orchard TJ, Wagener DK, Becker DJ, Cavender DE and Drash AL. The Pittsburgh insulin-dependent diabetes mellitus (IDDM) morbidity and mortality study. Mortality results. *Diabetes* 1984, 33(3): 271-276.

Eachus J, Williams M., Chan P, Smith GD, Grainge M, Donovan J and Frankel S. Deprivation and cause specific morbidity: evidence from the Somerset and Avon Survey of Health. *BMJ* 1996, 312: 287-292.

Fabré J, Balant LP, Dayer PG, Fox HM, Vernet AT. The kidney in maturity onset diabetes mellitus. A clinical study of 510 patients. *Kidney Int* 1982; 21: 730-8.

Fuller JH, Elford J, Goldblatt P and Adelstein AM. Diabetes mortality: new light on an underestimated public health problem. *Diabetologia* 1983, 24: 336-341.

Gerard K, Donaldson C and Maynard AK. The cost of diabetes. *Diabetic Medicine* 1989, 6: 164-170.

Gray A, Fenn P and McGuire A. The cost of insulin-dependent diabetes mellitus (IDDM) in England and Wales. *Diabetic Medicine* 1995, 12: 1068-1076.

Green A and Hougaard. Epidemiological studies of diabetes mellitus in Denmark: clinical characteristics of insulin-treated diabetes. *Diabetologia* 1983, 25:231-234.

Henderson J, Goldacre MJ, Griffith M and Simmons H. Recording of deaths in hospital information systems. Implications for audit and outcome studies. *Journal of Epidemiology & Community Health* 1992, 46(3): 297-299.

Henderson M. "Ailing inner London tops Europe's sick list". *The Times* 1998, 21/9/1998.4

Hull D. Cost Modelling - *MSc thesis in Medical Informatics* 1997.

Irving D. Identification of underprivileged areas (Technical Paper). *London: School of Economics*, 1983.

Jarman B. Identification of underprivileged areas. *B MJ* 1983, 286: 1705-9.

Jarman B. Deprivation indices. *BMJ* 1991, 303: 323.

Jarman B. Underprivileged areas; validation and distribution of scores. *BMJ* 1984; 289: 1587-1592.

Jarrett RJ, McCartney P and Keen H. The Bedford survey: ten year mortality rates in newly diagnosed diabetics, borderline diabetics and normoglycaemic controls and Risk indices for coronary heart disease in borderline diabetics. *Diabetologia* 1982, 22:79-84.

Jarrett RJ, Viberti GC, Argyropoulos A, Hill RD, Mahmud U, Murrells TJ. Microalbuminuria predicts mortality in non-insulin-dependent diabetics. *Diabetic Medicine* 1984, 1: 17-19.

Jarrett RJ. Type 2 (non-insulin-dependent) diabetes mellitus and coronary heart disease - chicken, egg or neither? *Diabetologia* 1984, 26: 99-102.

Kalbfleishch JD and Prentice RD. The statistical analysis of failure time data. *Wiley* 1980.

Kannel WB and McGee DL. Diabetes and glucose tolerance as risk factors for cardiovascular disease: the Framingham study. *Diabetes Care* 1979, 2: 2, 120-125.

Kelly WF, Mahmood R, Kelly MJ, Turner S, Elliot K. Influence of social deprivation on illness in diabetic patients. *Br Med J* 1993; 307: 115-116.

Kelly WF, Mahmood R, Turner S, Elliott K. Geographical mapping of diabetic patients from the deprived inner city shows less insulin therapy and more hyperglycaemia. *Diabetic Medicine* 1994, 11: 344-348.

Klein R, Moss SE, Klein BEK, DeMets DL. Relation of ocular and systemic factors to survival in diabetes. *Arch Int Med* 1989, 149: 266-272.

Kleinman JC, Donahue RP, Harris MI, Finucane FF, Madans JH and Brock DB. Mortality among diabetics in a national sample. *American Journal of Epidemiology* 1988, 128: 2 389-401.

Knuiman MW, Welborn TA and Whittal DE. An analysis of excess mortality rates for persons with non-insulin-dependent diabetes mellitus in Western Australia using the Cox proportional hazards regression model. *American Journal of Epidemiology* 1992, 135 (6) : 638-648.

Lambeth Southwark & Lewisham Health Authority Executive Summary. Annual Report of the Director of Public Health 1995/1996.

Leese B. The costs of diabetes and its complications. *Social Med* 1992, 35(10): 1030-1310.
Main J, Main P. Jarman index. *BMJ* 1991, 302: 850 (6 April)

Main J, Main P. Jarman index. *BMJ* 1991; 302: 850 (6 April)

Marmot MG, McDowall ME. Mortality decline and widening social inequalities. *Lancet* 1986, 2: 274 - 276.

McCarty D and Zimmet P. Diabetes 1994 to 2010: Global estimates and projections. International Diabetes Institute, Melbourne, Australia - a WHO collaborating centre for diabetes mellitus. 1994, *Bayer AG, 51368 Leverkusen, Germany, International Diabetes Institute, Melbourne, Australia.*

Morris R and Carstairs V. Which deprivation? A comparison of selected deprivation indexes. *Journal of Public Health Medicine* 1991, 13 (4) : 318-326.

Morrish NJ, Stevens LK, Fuller JH, Keen H and Jarrett RJ. Incidence of macrovascular disease in diabetes mellitus: the London cohort of the WHO Multinational Study of Vascular Disease in Diabetics. *Diabetologia* 1991, 34: 584-589.

Morrish NJ, Stevens LK, Head J, Fuller JH, Jarrett RJ and Keen H. A prospective study of mortality among middle-aged diabetic patients (the London cohort of the WHO Multinational Study of Vascular Disease in Diabetics) II: associated risk factors. *Diabetologia* 1990, 33: 542-548.

Mulrow C, Bailey S, Sönksen PH and Slavin BM. Evaluation an audiovisual diabetes education program: negative result of a randomised trial of patients with non-insulin dependent diabetes. *Journal of General Internal Medicine* 1987, 2:215-219.

Neal RD, Heywood PL and Morley S. Real word data retrieval and validation of consultation data from four general practices. *Family Practice* 1996, 13 (5) : 455-61.

Newsletter of the Juvenile Diabetes Foundation (UK). *Diabetes breakthrough* 1996, 9: 3.

NHS Number & Tracing Services (AR) programme. The NHS Number Tracing Service: guidance on the management of routine tracing and trace outcomes. *IMG publication* 1997, C3219.

NHSME 1993b. Computerisation in GP practices; 1993 Survey. *London, Department of Health and the Central Office of Information, London, NHSME.*

Nyström L, Wall S, Wibell L and the Diabetes Incidence Study in Sweden (DISS) Group. Mortality of all incident cases of diabetes mellitus in Sweden diagnosed 1973-1987 at age 15-34 years. *Diabetic Medicine* 1992, 9: 422-427.

Ochi JW, Melton LJ, Palumbo PJ and Chu CP. A population-based study of diabetes mortality. *Diabetes Care* 1985, 8 (3): 224-229.

Office of Population Censuses & Surveys. Guidance notes on applications to use the NHS Central Registers for Research. 1-7.

Pan WH, Cedres LB, Liu K, Dyer A, Schoenberger JA, Shekelle RB, Stamler R, Smith D, Collette P and Stamler J. Relationship of clinical diabetes and asymptomatic hyperglycemia to risk of coronary heart disease mortality in men and women. *American Journal of Epidemiology* 1985, 123: 504-516.

Panzram G, Zabel-Langhennig R. Prognosis of diabetes mellitus in a geographically defined population. *Diabetologia* 1981, 20:587-591.

Panzram G. Mortality and survival in Type 2 (non-insulin-dependent) diabetes mellitus. *Diabetologia* 1987, 30: 123-131.

Philip W, James T, Nelson M, Ralph A and Leather S. The contribution of nutrition to inequalities in health. *BMJ* 1997, 314:1545-1549.

Published East Sussex FHSA, Diabetes care management. *First Health* Summer 1994,6-7.

Ratner RE. Long-term health care outcomes in diabetes. *Endocrinology and Metabolism Clinics of North America* 1997, 26(3) 487-498.

Rossing P, Hogaard P Borch-Johnsen K, Parving HH. Predictors of mortality in insulin dependent diabetes: 10 year observational follow-up study. *BMJ* 1996, 313: 779-784.

Sönksen PH, Harris A and Jarman B. Perceived strategic and practical problems in the use of information technology for quality improvement in diabetes care in the United Kingdom. *Diabetes & Metabolism* (Paris) 1993, 19, 80-88.

Sasaki A, Uehara M, Horiuchi N and Hasagawa K. A long-term follow-up study of Japanese diabetic patients: mortality and causes of death. *Diabetologia* 1983, 25: 309-312.

Schmitz A and Vaeth M. Microalbuminuria: a major risk factor in non-insulin-dependent diabetes. a 10 year follow-up study of 503 patients. *Diabetic Medicine* 1988, 5: 126-134.

Shenfield GM, Elton RA, Bhalla IP, Duncan LSP. Diabetic mortality in Edinburgh. *Diabetes Metabolism* 1979, 5: 149-158.

Shillitoe RW. Psychology and Diabetes - Psychosocial factors in management and control. *Chapman and Hall, London* 1988, 226-227.

Sprafka JM, Pankow J, McGovern PG and French LR. Mortality among Type 2 diabetic individuals and associated risk factors: the Three City Study. *Diabetic Medicine* 1993, 10: 627-632.

Stengard JH, Tuomilehto J, Pekkanen J, Kivinen P, Kaaraslo E, Nissinen A and Karvonen MJ. Diabetes mellitus, impaired glucose tolerance and mortality among elderly men: the Finnish cohorts of The Seven 'countries Study. *Diabetologia* 1992, 35: 760-765.

The newsletter for ONS' medical research services. The Researcher. *Office for National Statistics* 1996, 6: 2.

Thompson EJ. Statement of policies on confidentiality and security of personal data. *Issued by OPCS* 1993, 2.

Townsend P, Davidson N, Whitehead M, Eds. Inequalities in health. *The Black Report and the Health Divide. London: Penguin Books* 1992.

Townsend P, Phillimore P and Beattie A. Inequalities in health in the Northern region. *Newcastle Upon Tyne: North Regional Health Authority and University of Bristol* 1986.

Vaughan NJ and Home PD. The UK Diabetes Dataset: a standard for information exchange Diabetes Audit Working Group of the Research Unit of the Royal College of Physicians. British Diabetic Association. *Diabetic Medicine* 1995, 12(8): 717-22.

Vaughan NJ. Measuring the outcomes of diabetes care. Audit Working Group of the Research Unit of the Royal College of Physician and British Diabetic Association. [Meeting Report] *Diabetic Medicine* 1994, 11(4): 418-23.

Vaughan NJA, N. Hopkinson, V.A. Chishty. DIALOG: co-ordination of the annual review process through a district diabetes register linked to the FHSA database. *Diabetic Medicine* 1996, 13: 182-188.

Vaughan NJA, Shaw M, Boer F, Billett D, Martin C. Creation of a District Diabetes Register using the DIALOG system. *Diabetic Medicine* 1996, 13: 175-181.

Viberti GC, Hill RD, Jarrett RJ, Argyropoulos A, Mahmud U, Keen H. Microalbuminuria as a predictor of nephropathy in insulin dependent diabetes mellitus. *Lancet* 1982, 1: 1430-1432.

Walters DP, Gatling W, Houston AC, Mullee MA, Julious SA and Hill RD. Mortality in diabetic subjects: an eleven-year follow-up of a community-based population. *Diabetic Medicine* 1994, 11: 968-973.

Wang SL Cause-specific mortality rates in diabetic patients: WHO multinational study. *PhD thesis* 1996.

Ward JD. Wealthy means healthy: diabetes and social deprivation. *Diabetic Medicine* 1994, 11: 334-335.

Warner DC, McCandless RR, DeNino LA, Cornell JE, Pugh JA and Marsh GM. Costs of diabetes in Texas, 1992. *Diabetes Care* 1996, 19(12) :1416-1419.

Waugh NR, Dallas JH, Jung RT and Newton RW. Mortality in a cohort of diabetic patients. Cause and relative risks. *Diabetologia* 1989, 32(2): 103-104.

Wei LJ. The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine* 1992, 11: 1871-1879.

Weng C, Carey S, Coppini DV, Sönksen PH. The feasibility study on record linkage between a hospital diabetes database and other relevant information systems within the National Health Service. Diabetes, *Nutrition & Metabolism Clinical* 1998, 11(1) P34.

Weng C, Coppini DV AND Sönksen PH. Are geographic and social factors related to increased morbidity and mortality rates in diabetic patients? *Diabetic Medicine* 1997, 14 [Suppl 4].

Weng C, Coppini D, Mozzaka N, Sönksen PH. Deaths related to diabetic foot problems as an estimate of mortality associated with peripheral neuropathy. *Diabetic Medicine* 1996, 13 [Suppl 7] P79 (Abstract).

Weng C, Coppini DV and Sönksen PH. Linking a hospital diabetes database and the National Health Service Central Register: a way to establish accurate mortality and movement data. *Diabetic Medicine*, 1997, 14: 877-883.

Will JC and Connell FA. The preventability of 'premature mortality': an investigation of early diabetes deaths. *American Journal of Public Health* 1988, 78: 831 - 833.

William CD, Harvey FE, Sönksen PH. A role for computers in diabetes care? *Diabetic Medicine* 1988, 5: 177-199.

William D. Alexander and South East Thames Diabetes Physicians Group. Diabetes care in a UK health region: activity, facilities and costs. *Diabetic Medicine* 1988, 5: 577-581.

Wilson AE and Home PD. A dataset to allow exchange of information for monitoring continuing diabetes care. The Diabetes Audit Working Group [see comments]. *Diabetic Medicine* 1993, 10(4): 378-90.

Wong JSK, Pearson DWM, Murchison LE, Williams MJ and Narayan V. Mortality in diabetes mellitus: experience of a geographically defined population. *Diabetic Medicine* 1991, 8: 135-139.

Woodward M, Tunstall-Pedoe M, Smith W, Tavendale R. Smoking characteristics and Inhalation Biochemistry in the Scottish Population. *J Clin Epidemiol* 1991, 44: 1405-1410.

Young MJ, Breddy JL, Veves A, Boulton AJM. The prediction of diabetic neuropathic foot ulceration using vibration perception thresholds: a prospective study. *Diabetes Care* 1994, 17: 557-560.

Young RJ. Assessing Diabetes Care Services. *Draft material from an RCP, RCGP, BDA working group* 1999.

Appendix A Feasibility study on records linkage between "Diabeta" and all relevant information systems

System name	Clinical data item	Data completeness	Range of population recorded in the system	Data used for matching records	Type of matching	Availability of information on costing
Accident + Emergency own computer system (STH)	• Date of attendance • Problem (free text) • Outcome (destination)	Good	Patients who have attended Accident and Emergency in STH since 1992	• Hospital number • Patients name • Date of birth • Address	Manual	Cannot identify cost per visit, because the payment is based on a contractual package with LSL Health Authority
CMMS (STH & GUY)	• Date episode started • Date episode ended • Consultant code Diagnosis (ICD 9/ICD 10) • Operation procedures (OPCS code) • Speciality code	Good HA audit the quality of data supplied by each hospital every half year.	Patient who had re/admissions in STH&GUY after 1990. Patients had an out-patient attendance in STH &GUY after 1997/1995	• Hospital number	Automatic matching through hospital network	The cost for each episode can be calculated based on a speciality code from CMMS
Result Reporting System (RRS) (STH & GUY)	• Date of laboratory test • Report title • Results (free text) • Abnormal	Good	Patients who had chemical pathology, microbiology and haemophilia test within half year. Cardiology and Radiology are now linked to RRS (Dec 97).	• Hospital number	Automatic matching through hospital network	It is included in the payment on consultant episode or contractual package for out-patients. Nevertheless standard costs are available, as is archived data
Contract Minimum Data Set (CMDS) (LSL HA)	• Provider number (Hospital number) • Date of episode started • Date of episode ended • Consultant code Diagnosis (ICD 9/10) Operation procedures (OPCS code)	Good	All residents living within LSL area who have been admitted to a hospital since 1987	• Postcode • Sex • Date of birth New NHS number (only available for records since April 1997)	Automatic matching. Data transfer could be on an EDI format through the NHSnet in the future	The cost for each episode can be calculated based on speciality code from CMMS

Appendix A Feasibility study on records linkage between "Diabeta" and all relevant information systems

System name	Clinical data item	Data completeness	Range of population recorded in the system	Data used for matching records	Type of matching	Availability of information on costing
Renal system ("Proton") (STH & GUY)	• Clinical details of ESRF, Dialysis and transplant.	Reasonable	Patients who are admitted for dialysis and transplant in STH & GUY.	• Hospital number • Patient's name • Date of birth • Address	Manually matched	Cost per treatment (package price). e.g. £1800 per month for dialysis.
General Practice computer systems (E&W)	• Treatment (Read code) • Repeat prescription • Hospital referral • Other specialities referral	About 95% of GP are using computers. The quality of data recorded on the system is different in each practice.	Residents who registered with a local GP.	• New NHS number • Patient's name • Address • Date of birth • Practice code (or postcode) • GP details	Sending a letter to each GP with requirement of patient's data.	There is no formal price per consultation. The prescription can be calculated based on the GP's computer systems.
"Comcare" (Lambeth Health Care)	• Type of visit • Services provide (local code)	Reasonable	Residents living within Lambeth area who had a home visit by district nurses since 1996	• Patient's name • Date of birth • Address	Manual matching. Data transfer by a floppy disk.	Defined cost for per District Nurse visit is available.
Social Services system (Lewisham Council)	• Type of disability • Type of services provided	Unknown	Residents living within Lewisham area who applied for disability benefits.	• Patient's name • Date of birth • Address	Automatic matching. Data transfer by floppy disks.	The cost is different for different type of services
Prescription Pricing Authority (PPA) in England	• GP's code • Prescription details (BMF code)	Good	Residents in England & Wales who purchase drugs from pharmacies/chemists based on a GP prescription.	Impossible to match on individual patients.	'flagging' study	The costs of drugs can be identified on groups of patients (see Section Four)

Appendix A Feasibility study on records linkage between "Diabeta" and all relevant information systems

System name	Number of records on the database	Frequency of updating (entering) data	Linked with other systems
Accident + Emergency own computer system (STH)	200~250 attendance per day, 80,000 records per year.	Daily entry	None
CMMS (STH & GUY)	10,000 episodes per year (in-patients). Half million per year out-patients.	New out-patients are added twice a week from COPS, Every weekend night data is transferred from HMS or PAS to CMMS.	Linked with hospital systems (HMS, COPS, PAS) and also NWCS * PAS will replace HMS and COPS.
Result Resource Report System (RRS) (STH & GUY)	Unknown	Daily entry	Linked with the PAS and other clinical examination systems in St. Thomas' Hospital.
Contract Minimum Data Set (CMDS) (LSL HA)	1,000,000 records by 1997	Unknown	Linked with each hospital's CMMS system
Renal system (Proton) (STH & GUY)	1,550 records by 1997	Daily entry	Linked with the RRS
General Practice own computer systems (E&W)	Different in each general practice. Each GP can only have 2000~3000 patients on his/her list.	Daily entry for detail of consultation, the letter received from other provider are entered later by medical staff.	Linked with patients' Registration System (EXETER) in a local Health Authority. Some of GP computer systems are linked with hospital lab. report system.
"Comcare" (Lambeth Health Care)	2,370 new attendance every half year.	Every quarter of month	None
Social Services system (Lewisham Council)	Unknown	Daily entry	Unknown
Prescription Pricing Authority (PPA) system in England	Unknown	Monthly	PPA collects prescriptions, which are prescribed by all GPs in England, from all pharmacies and chemists in the community for monitoring drug dispensing in the NHS.

Appendix B A copy of a death certificate

REGISTRATION DISTRICT : WANDSWORTH 256
REGISTRATION SUB DISTRICT : WANDSWORTH 1C
REGISTER NUMBER : C103A
ENTRY NUMBER : 177
ADMINISTRATIVE AREA : LB OF WANDSWORTH

1. DATE AND PLACE OF DEATH : TWENTY-FIRST SEPTEMBER 1998
ST GEORGE'S HOSPITAL, TOOTING
2. NAME AND SURNAME :
3. SEX : MALE
4. MAIDEN NAME : ----------
5. DATE & PLACE OF BIRTH :
YORK
6. OCCUPATION & USUAL ADDRESS: CLERK (RETIRED)

7.(A) NAME OF INFORMANT : PAUL STUBBS
(B) QUALIFICATION : SON PRESENT AT THE DEATH
(C) USUAL ADDRESS : DOVE WORKSHOPS, BARTON ROAD, BUTLEIGH, GLASTONBURY, SOMERSET
8. CAUSE OF DEATH : 1A. METASTATIC COLONIC ADENO CARCINOMA
B.
C.
11. NON INSULIN DEPENDENT DIABETES MELLITUS
.

CERTIFIED BY S. EVANS M.B.
(IV) ORIGINAL UNDERLYING ICD9: 1539
(IX) MULTIPLE ORIGINAL ICD9 : 1539
2500
9. INFORMANTS SIGNATURE : P. STUBBS
10.DATE OF REGISTRATION : TWENTY-SECOND SEPTEMBER 1998
11.SIGNATURE : M.A.ROSSALL REGISTRAR
O NHS NUMBER :
LINENO9 : 1
6
CODROW : 1
10
WIGGLESWORTH CODE : NOT PRESENT

492 672 3743

Appendix C1 A copy of GP referral letter (printed out from a computer)

THE GRANTHAM CENTRE Beckett House, Grantham Road, SW9 9DL

Telephone: 0171-733 6191

Fax: 0171-737 2870

S S Wickremesinghe MBBS FRCS
Elizabeth P McGinn BA DRCOG MRCGP
Stephen M Bell MBBS BChD
Alison J Johnston MB BCh MRCGP DRCOG
Marie J Crocombe MBBS DRCOG

Dr. C. Lowy,
Consultant in Endocrinology,
St. Thomas' Hospital

5th January 1995

Dear Dr. Lowy,

RE: [illegible] DOB: 25/04/19 0
[illegible]
73 26 06

You wrote to me regarding the above named patient on the 11th July as he failed to keep two follow-up appointments. You were wondering whether he had moved away and this is not the case and the address you have got on his notes is the correct one. I saw him today as his blood sugar was 26.7 and his HBA1 was 16.4, serum creatinine 99. On talking to him I note that he has not been taking his Metformin as prescribed, as such I have asked him to take his Metformin in addition to his Glibenclamide which is 20 mg, daily. In view of the fact that he has diabetic retinopathy I would be grateful if you could keep an eye on him in your clinic.

Thank you for seeing him.

Yours sincerely,

Dr. S. S. Wickremesinghe

New Appt from DNA

Reminder 23/2/95

Appendix C2 A copy of GP referral letter (written by hand)

Dr. D.P.J. McCARTHY, T.D., M.B.Ch.B., F.R.A.C.G.P.
Dr. BRONWEN EVANS, M.R.C.G.P., D.C.H., D.R.C.O.G

353 CLAPHAM ROAD
LONDON SW9 9AE
071-622 2006

Diabetician
Diabetic Clinic
St Thomas Hospital

5.1.95.

Dear Professor

Re: ANVAR

12.1.41.

& ANSAR PICKERING Rd
SW4 6NF.

This woman has been referred
to the diabetic clinic as a newly
diagnosed diabetic
I should be grateful if you
could confirm the exact
position & write back to me
with
the blood sugar findings

She is 23. & should have been
started on glucose tolerance test.

She has no high blood
pressure & a history of chest
pain on Nifedipine. Not on BB

Thank you for your help

Appendix D The core dataset collected in "DIALOG"

214 to 215	Other Related Problems	YORN10	Y
216 to 217	Impotence	YORN10	N
218 to 219	Dialysis/Transplantation	YORN10	N
220 to 221	Angina	YORN10	N
222 to 223	Myocardinal Infarction Last Year	YORN10	N
224 to 225	Previous History of IHD	YORN10	N
226 to 227	Stroke Last Year	YORN10	N
228 to 229	Previous History of Cerebrovascular Disease	YORN10	N
230 to 231	Hypoglycaemia - Hospital Treated	-1 or 0 to 20	N
232 to 233	Hypoglycaemia - GP Treated	-1 or 0 to 20	N
234 to 235	Injection Sites Abnormal	YORN10	N
236 to 237	Risk Factors	YORN10	Y
238 to 239	Smoking (cigarettes per day or 1 = Yes)	-1 or 0 to 99	N
240 to 241	Alcohol (units per week, 1 unit = 10g)	-1 or 0 to 99	N
242 to 243	Number of admissions for foot problems	-1 or 0 to 20	N
244 to 245	Number of admissions for Hypoglycaemic Emergencies	-1 or 0 to 20	N
246 to 247	Number of admissions for cardiac problems	-1 or 0 to 20	N
248 to 250	Number of admission days	-1 or 0 to 365	N
251 to 253	Sick Days	-1 or 0 to 365	N
254 to 255	Appointments Attended last year	-1 or 0 to 20	N
256 to 257	Appointments Missed last year	-1 or 0 to 20	N
258 to 259	Dietitian Attended	YORN10	N
260 to 261	General Depressed Score	-1 or 0 to 18	N
262 to 263	Diabetes Depressed Score	-1 or 0 to 18	N
264 to 278	Satisfaction Score	See notes	N
279 to 280	Chiropodist Attended	YORN10	N

Lists of Values

YORN10

These fields are simple Yes or No answers. However, some systems already hold this data in the format 0 or 1 therefore Dialog will accept the data in either format.

null (blank or 2 spaces)	-1	These values mean that no data has been recorded.
N	0	This means NO or logical false.
Y	1	This means YES or logical true.

E.g. The values 'Y' or '1' for 'Pin Prick Abnormal - Right' means YES the pin prick for the right foot was abnormal.

E.g. The values 'N' or '0' for 'Chiropodist Attended' means NO the Chiropodist did not attend.

<u>For the mandatory items the values 'null' and '-1' are not valid.</u>

Self Monitoring Type

-1 or null (2 spaces)	Unknown/Not Recorded
0	None
1	Urine
2	Blood
3	Both

Visual Acuity

These are expressed as 'Snellian as decimal.'
-1 or null = not recorded
Valid values are 6.05, 6.06, 6.09, 6.12, 6.18, 6.24, 6.36, 6.60

Satisfaction Score

Each of the fifteen digits represents a question, each being scored 0, 1, 2 or 3. A missing result is expressed as '-'.

Appendix D The core dataset collected in "DIALOG"

DIALOG - INFORMATION FOR ANNUAL REVIEW DISKS

Format for data for annual reviews. This can be read in from a DOS formatted ASCII file on disk and then converted into a unix file. Alternatively, Dialog will read in information direct from a unix file as long as it is in the order specified below.

Characters	Field	Values	Mandatory?
1 to 14	NHS Number	Anything	Y
15 to 22	Date of Review	DDMMYYYY	Y
23 to 29	Date of Next Review	MM/YYYY	N
30 to 35	Lead Clinician	National 6 figure number	Y
36 to 46	Site Code	National Code	N
47 to 48	Patient Attended	YORN10	Y
49 to 50	Gestational Diabetes	YORN10	N
51 to 52	Patient Pregnant	YORN10	N
53 to 54	Insulin Treated	YORN10	N
55 to 56	Tablet Treated	YORN10	N
57 to 58	Diet Treated	YORN10	N
59 to 63	Weight (kilogrammes)	-1 or 0 to 300.0	N
64 to 68	Height (metres)	-1 or 0 to 2.500	N
69 to 71	Blood Pressure - systolic (mmHg)	-1 or 0 to 999	N
72 to 74	Blood Pressure - diastolic (mmHg)	-1 or 0 to 999	N
75 to 76	Blood Pressure Treatment	YORN10	N
77 to 78	Angina Treatment	YORN10	N
79 to 80	Cardiac Failure Treatment	YORN10	N
81 to 82	Anti Lipid Treatment	YORN10	N
83 to 84	Self Monitoring Type	See codes	N
85 to 86	Measurements Taken	YORN10	Y
87 to 90	Visual Acuity (Right)	See codes	N
91 to 94	Visual Acuity (Left)	See codes	N
95 to 98	Fasting Glucose (mmol/l)	-1 or 0 to 90.0	N
99 to 102	Hba1/Hba1c (%Hb)	-1 or 0 to 90.0	N
103 to 107	Fructosamine (mmol/l)	-1 or 0 to 900.0	N
108 to 109	Albustix positive	YORN10	N
110 to 115	Albumin/Creatinine Ratio (mg/mmol)	-1 or 0 to 9999.0	N
116 to 121	Albumin Excretion Rate (ug/min)	-1 or 0 to 9999.0	N
122 to 125	Creatinine (umol/l)	-1 or 0 to 2000	N
126 to 129	Cholesterol (mmol/l)	-1 or 0 to 30.0	N
130 to 133	Triglycerides	-1 or 0 to 99.0	N
134 to 135	Eyes Examined	YORN10	Y
136 to 137	Optician/Optometrist Attended	YORN10	N
138 to 139	Ophthalmologist Attended	YORN10	N
140 to 141	Retinal Photography Performed	YORN10	N
142 to 143	Ophthalmoscopy Performed	YORN10	N
144 to 145	Retinal Summary Abnormal	YORN10	N
146 to 147	Advanced Retinal Disease	YORN10	N
148 to 149	Retina Not Visualised - Right	YORN10	N
150 to 151	Retina Not Visualised - Left	YORN10	N
152 to 153	Cataract Present - Right	YORN10	N
154 to 155	Cataract Present - Left	YORN10	N
156 to 157	Cataract Extracted - Right	YORN10	N
158 to 159	Cataract Extracted - Left	YORN10	N
160 to 161	Background Retinopathy - Right	YORN10	N
162 to 163	Background Retinopathy - Left	YORN10	N
164 to 165	Maculopathy - Right	YORN10	N
166 to 167	Maculopathy - Left	YORN10	N
168 to 169	Pre-proliferative Retinopathy - Right	YORN10	N
170 to 171	Pre-proliferative Retinopathy - Left	YORN10	N
172 to 173	Proliferative Retinopathy - Right	YORN10	N
174 to 175	Proliferative Retinopathy - Left	YORN10	N
176 to 177	Laser/Vitrectomy - Right	YORN10	N
178 to 179	Laser/Vitrectomy - Left	YORN10	N
180 to 181	Feet Examined	YORN10	Y
182 to 183	Pin Prick Abnormal - Right	YORN10	N
184 to 185	Pin Prick Abnormal - Left	YORN10	N
186 to 187	Vibration Sense Abnormal - Right	YORN10	N
188 to 189	Vibration Sense Abnormal - Left	YORN10	N
190 to 191	Symptomatic Neuropathy	YORN10	N
192 to 193	Foot Ulcer Present - Right	YORN10	N
194 to 195	Foot Ulcer Present - Left	YORN10	N
196 to 197	Absent Foot Pulses - Right	YORN10	N
198 to 199	Absent Foot Pulses - Left	YORN10	N
200 to 202	Claudication - Right	YORN10	N
202 to 203	Claudication - Left	YORN10	N
204 to 205	Amputation Toe/Forefoot - Right	YORN10	N
206 to 207	Amputation Toe/Forefoot - Left	YORN10	N
208 to 209	Amputation Leg - Right	YORN10	N
210 to 211	Amputation Leg - Left	YORN10	N
212 to 213	Previous History of Peripheral Vascular Disease	YORN10	N

Linking a Hospital Diabetes Database and the National Health Service Central Register: a Way to Establish Accurate Mortality and Movement Data

C. Weng*, D.V. Coppini, P.H. Sönksen

Department of Diabetes, Endocrinology and Metabolic Medicine, United Medical and Dental School of Guy's and St Thomas' Hospitals, St Thomas' Hospital, London SE1 7EH, UK

We have established a records linkage between 'Diabeta' (the computerized clinical records system in the Diabetes Unit of St Thomas' Hospital) and the National Health Services Central Register (NHSCR) of the United Kingdom. Over 7000 diabetic patient records have been collected since 1973. Demographic data on all diabetic patients were retrieved and submitted to the NHSCR via a floppy disk. A matching system (automatic or manual) was used by the NHSCR to identify deceased patients and the most recent demographic data was provided on patients alive. This linkage resulted in an update of 91 % of records in Diabeta. The findings of the update included:

(1) 86 % of diabetic patient's death had not been notified to the hospital and were not recorded on Diabeta. Mortality can now be assessed accurately as an outcome measure in our diabetic population.

(2) Provision of the NHS number to Diabeta, as before it was not available on many patients seen in the hospital. The NHS number is a key patient identifier which can be used to exchange information within the NHS-wide network.

(3) Diabetes was recorded as a cause of death in only 36% of death certificates. Analyses of death certificates alone must thus give poor information about mortality in diabetes.

(4) Geographical location of patients on the database was updated, enabling tracing of patients for long-term studies and analyses of movement.

***Diabet. Med.* 14: 877–883 (1997)**

No of Figures: 4. No of Tables: 0. No of Refs: 17

KEY WORDS **Diabetes database Diabeta Medical records linkage NHS Central Register Mortality Demographic data Prospective studies**

Received 6 February 1997; accepted 2 May 1997

Introduction

The St Vincent Declaration[1] has raised awareness about the importance of undertaking comprehensive and systematic actions for preventing the complications of diabetes mellitus. The declaration discussed the importance of information technology (IT) in supporting the management of patients with diabetes. Two important roles of IT are in coping with increasing geographical movement of patients and the needs of more specialist oriented health care. There is increasing demand for better communication between computerized systems handling data on patients with diabetes in various health care sectors and other establishments (e.g. health authorities, insurance companies). Information systems in each area need to be completed and then linked;[2] the challenge is to accomplish this throughout Europe.

In this study, a records linkage between Diabeta, the computerized diabetes clinical record systems of the Diabetes Unit at St Thomas' Hospital, and some Health Authorities is assessed.

In 1980, a real-time computerized clinical records system (Diabeta) was implemented to improve the quality of care of diabetic outpatients at St Thomas' Hospital. At present, there are about 10 000 diabetic patients recorded.[3] About 400 new cases are entered into Diabeta each year. The accuracy of the information on our patients' vital status (alive/deceased) has always been acknowledged to be poor, as notification of death to the hospital is notoriously incomplete. Likewise we know that maintaining accurate demographic data on patients in Diabeta is difficult since within an inner city as many as 30 % of people change some details each year. These changes cannot be updated, unless the patients, their relatives or their GP notifies us of them. Thus accurate

* Correspondence to: Miss Chao Weng, Division of Medicine, North Wing, 4th Floor, St Thomas' Hospital, London SE1 7EH, UK.
Sponsors: The British Council; The Directorate of Diabetes and Endocrinology, Guy's and St Thomas' Hospital Trusts

CCC 0742–3071/97/100877–07$17.50

analysis of mortality and our ability to perform longitudinal studies in our diabetes clinic population is impossible unless the patients' vital status is continually updated. Establishing a better communication system between Diabeta and another reliable information source would be a possible solution. The National Health Service Central Register (NHSCR) which is a division of the Office of Population Census and Surveys (OPCS) (now known as the Office for National Statistics (ONS)) is one such important source.

In order to assess the suitability of developing a records linkage between Diabeta and the NHSCR, three questions need to be answered:

1. To what extent could the NHSCR be used to update paients' vital status and demographic data in Diabeta?
2. What are the benefits and costs of developing such a linkage?
3. What methods can be used to ensure confidentiality on patient's information?

To answer these questions, a survey on the NHSCR system was carried out.

The NHSCR keeps updated demographic data on all persons registered with a GP or NHS doctor in England and Wales.[2] The NHSCR established a computerized registration system in 1991 and keeps records on all people known to be alive since 1 January 1991. Manual records go back much further.

Each patient's record in the NHSCR database contains the following demographic details: name, sex, date and place of birth, NHS number, FHSA code and date of registration, and if deceased: date and place of death. Home address, details of current GPs, and clinical information are not recorded.

The NHSCR database is linked 'on-line' with all Family Health Services Authorities (FHSAs; now known as Health Authorities)[4] in England and Wales. Each FHSA is responsible for registering individuals with a GP within its boundary and for notifying new registrations to the NHSCR. The NHSCR also routinely receives notifications from all Local Registrars of Births and Deaths and from other organizations (e.g. embarkation, long-stay psychiatric institutions, prisons, and the Armed Forces in England and Wales).

The NHSCR is used to distribute information between Family Health Services Authorities (FHSAs), in order to update people's registration status in each FHSA registration database. It is also responsible for informing the corresponding FHSA when its patients 'exit' from the NHS (such as death, embarkation, long-stay institutions, Army, etc.). The NHS number is recorded on each registration in the NHSCR. This is a key patient identifier and is used to exchange information between the FHSA registration database and the NHSCR database within the NHS network. The old record systems (manual cards, microfilms, and microfiches) were however also retained by the NHSCR. The population registration process within the NHS is illustrated in Figure 1.

The NHSCR database has been used to assist researchers to integrate patients' demographic data in their local database with that in the NHSCR. Prospective research on patients (flagging studies)[5] has been supported by the NHSCR system. For example, death certificates may be obtained using the 'flagging study'. Details on the 'flagging study' and its costs are described later in more detail. The new NHS number which will be introduced in the coming years will (where both co-exist) link to the old NHS number on the NHSCR database.

Patients and Methods

When this study was initiated, over 7000 diabetic patient records had been entered into Diabeta since it started in 1973. Patient demographic data (forename, surname, date of birth, gender, hospital number, last address, and date last seen in diabetic clinic) were retrieved from Diabeta.[5] These data were formatted on floppy disks according to the NHSCR regulations subject to special processing for data security. The floppy disks were mailed to the NHSCR to be updated. Records were automatically matched with the NHSCR database using the first four characters of each patient's surname and forename and with their date of birth. If all these items were consistent with analogous items on records in the NHSCR system, a match and link was considered to have been made. If there was more than one matching dataset in the NHSCR records, automatic matching alone was not successful. Such records would be skipped and the operator would carry on with the searching process by using other identification data (such as gender, address and last date seen in the diabetic clinic). Patients' addresses were not recorded in the NHSCR, but the NHSCR operator could use the patients' addresses for comparison with the FHSA area with which patients were registered. A 'trace model' was used to estimate the probability of both records belonging to the same person. In the case of poor quality identification data, and in patients who died or emigrated before 1991, record matching was done manually using the old record systems.

When patient records were matched and linked with records in the NHSCR, a unique study number was assigned to the patient records by the NHSCR. This study number, which was only disclosed to the authorized medical researchers in the NHSCR, had been entered (or 'flagged') into the NHSCR records. Therefore whenever a patient death occurred, a copy of the death certificate would be forwarded to us within a few weeks. Details of 'exits' from the NHS such as embarkation would also be notified. Meanwhile, the corrected discrepancies (name and date of birth) and patient identifiers (FHSA code, NHS number), which were not recorded in Diabeta, were also provided.[5] Transferring updated patients demographic data from the NHSCR to Diabeta is performed on a regular monthly basis. Meanwhile, new patient

Diabet. Med. 14: 877–883 (1997)

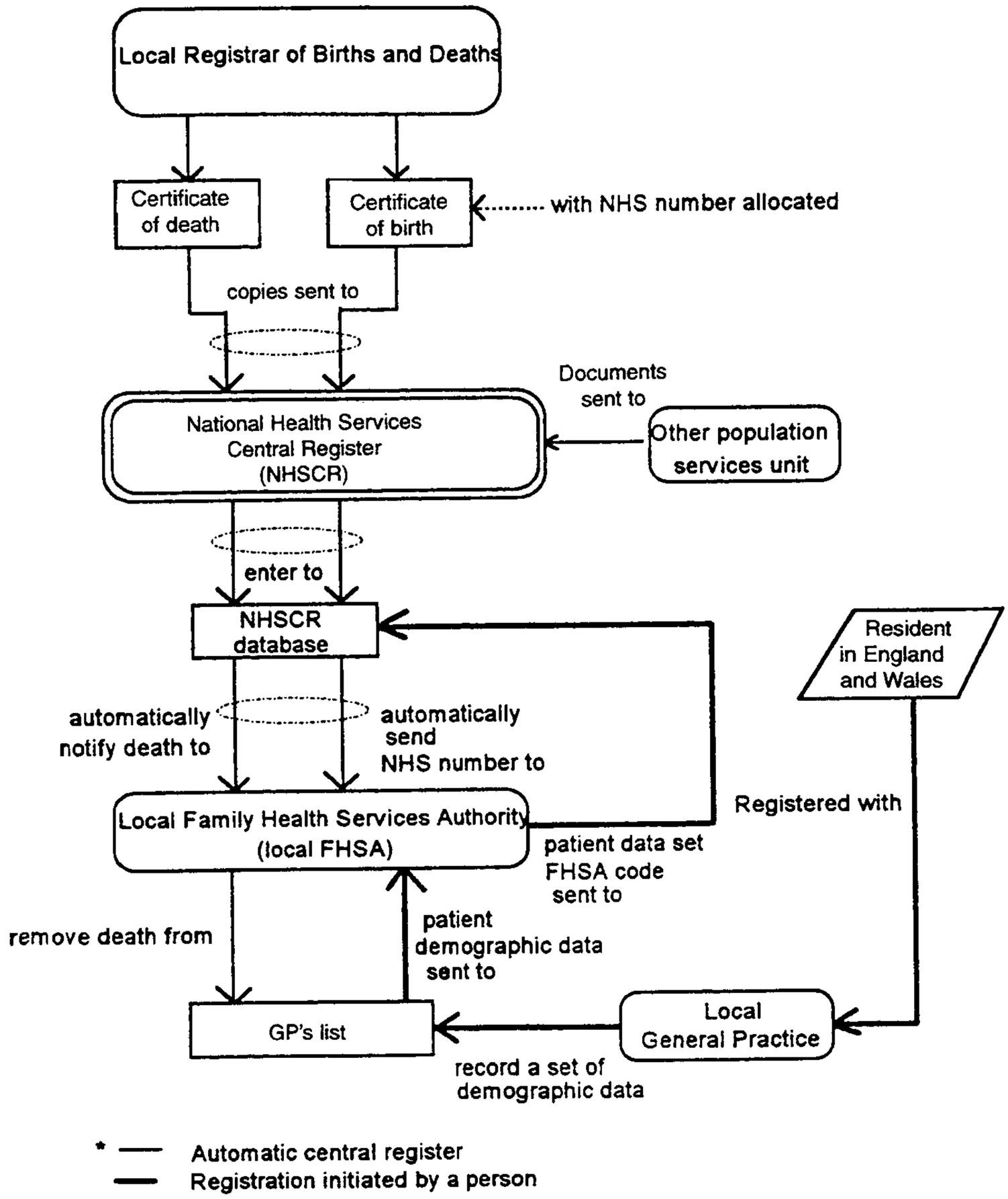


Figure 1. The population registration process within NHS

records are retrieved from Diabeta and sent to the NHSCR for prospective 'flagging' on a similar regular basis.

Handling of patient data on Diabeta and the NHSCR complied with the Data Protection Act.[6] In the case of record linkage between the NHSCR database and Diabeta, the NHSCR has issued security measures for ensuring transfer of authorized data.[7]

Results of updated records are currently provided in paper form. The NHSCR plans to provide these results electronically (on floppy disk) in the future. Data on floppy disks would then be subject to data security processing.[7] The protocol of the record linkage system between Diabeta and the NHSCR database is illustrated in Figure 2.

Results

The Possibility of Linkage Between Diabeta and the NHSCR

In total, 7542 diabetic records have been traced on the NHSCR. Of these, 6851 records (91 %) in Diabeta were successfully linked with records in the NHSCR (i.e. there was enough evidence to indicate valid matched ('linked') records belonged to the same patient.[8] Six hundred and fifty-five records (8.7 %) failed to match any existing records in the NHSCR. Thirty-six records (0.48 %) could only possibly be linked if complete patient identification information was supplied to the NHSCR (Figure 3).

Updating the Patient's Vital Status (Alive/Deceased) in Diabeta

The NHSCR provided information on the vital status of 91 % patients submitted. A total of 1670 deceased patient's records were traced by the NHSCR. In only 239 (14 %) of these patients had the death been notified to Diabeta.

Mention of 'Diabetes Mellitus' on Death Certificates

Copies of death certificates were supplied by the NHSCR. The *International Classification of Disease, Revision 9,*

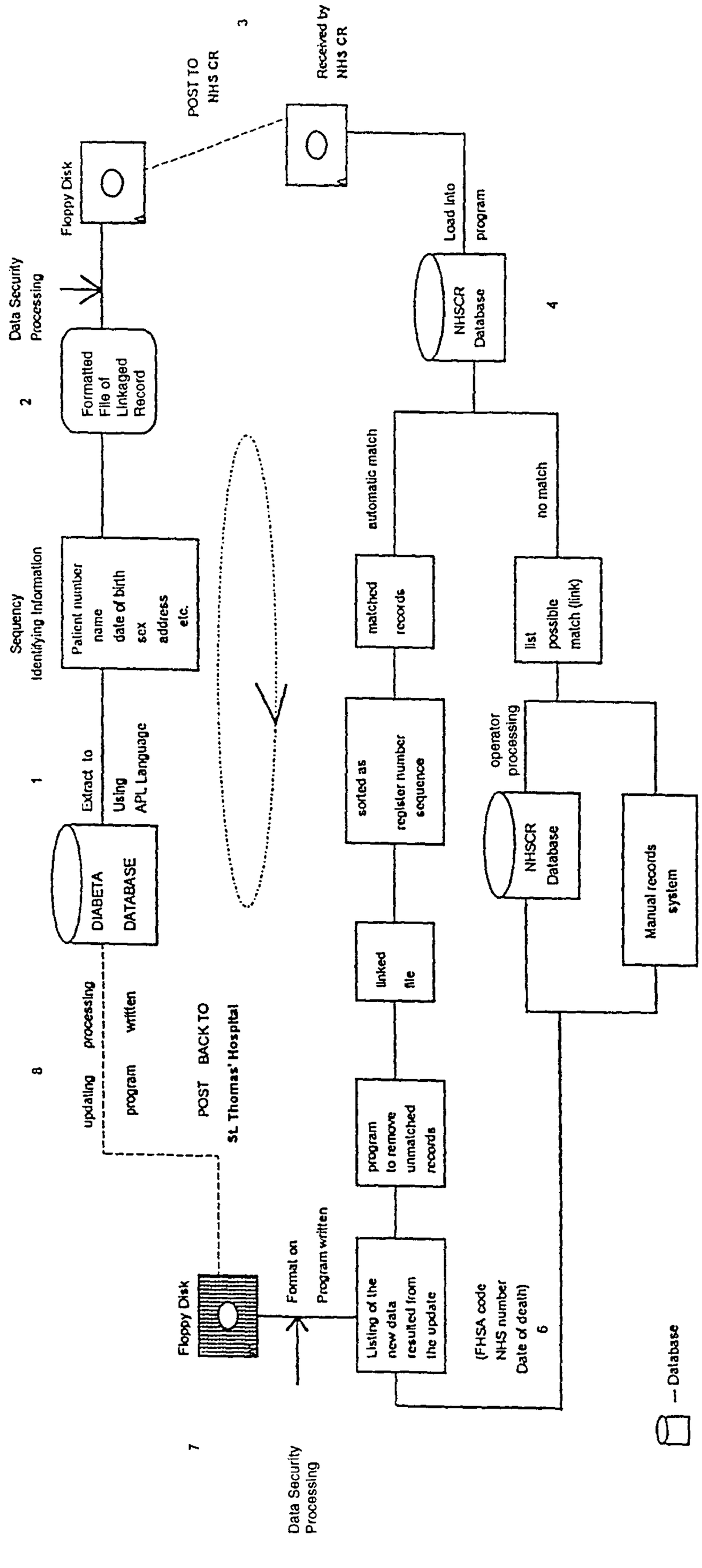


Figure 2. The records linkage system between 'Diabeta' Database and NHSCR

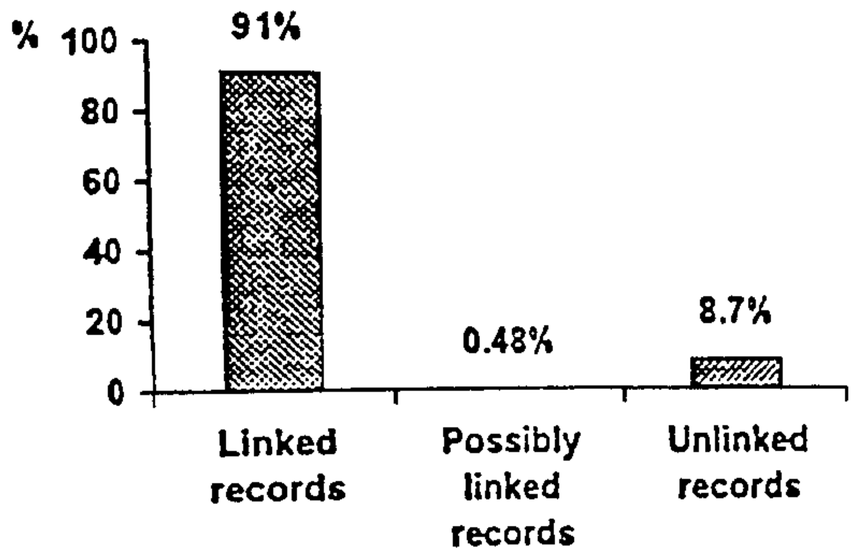


Figure 3. Linkage of records in 'Diabeta' with the NHSCR database

was used to code the cause of death by OPCS; diabetes mellitus corresponds to codes 250.0–250.9. 'Diabetes mellitus' was mentioned as an underlying or contributory cause of death in only 36 % of death certificates.[9]

Provision of FHSA Code for all Patients Alive Traced in the NHSCR

Nearly 94 % of our patients were still registered with an NHS doctor. For all these patients, codes of their current FHSA and the dates of registration were provided by the NHSCR, enabling us to trace patients' current GPs through the relevant FHSAs.

Three hundred and twenty (6 %) patients were no longer registered with the NHS. Their records would be marked by a different code (e.g. embarkation). In patients considered to 'exit' from the NHS, although their respective GP could not be traced, any deaths of such patients could be notified if they died in the UK.

NHS Number

Diabeta has a field for NHS number, but this was recorded < 1% of cases as the number has in the past been rarely used by GPs in their referral letters. The NHSCR has provided NHS numbers on all our patients which have now been added to the Diabeta database.

Numeric Distribution in Different Records Matching Procedures

Depending on the quality of patient data supplied to it, the NHSCR uses different procedures for matching records; on this occasion the quality of the data from Diabeta was sufficiently good to enable 42 % of the records to be matched automatically with the NHSCR database; 39% were matched on the NHSCR database 'semi-automatically' (requiring operator interference) and 19% were matched manually on the old record systems (Figure 4).

The cost of automatic matching of records was significantly lower than that of manual matching. If the NHS number had been recorded on Diabeta, the cost-savings for the electronic linkage between Diabeta and the NHSCR database would be significant. The average cost of 'flagging' events (e.g. death certificate, embarkation or changed FHSA registration) on the NHSCR database and retrospectively or prospectively notifying changes in patients' demographic or NHS registration status was between £1 and £2 per patient. This is a 'one off' charge and is not expensive considering the importance of having an accurate and complete database which does not send follow-up appointments to people who have died and can support meaningful epidemiological studies and reliable clinical audits. These costs can be even further reduced if patient records on Diabeta can be automatically matched with those of the NHSCR. Recently, better quality and quantity of patient identification information on Diabeta have led to more advanced and effective matching techniques with fewer costs.

Discussion

We have shown that the majority of records in our computerized patient record system Diabeta can be successfully traced and flagged on the NHSCR database. As a result, we discovered that the vast majority of deaths in our patients was not recorded by conventional means. Mortality analysis in our diabetic patient cohort would have been meaningless without the NHSCR link. We also identified that diabetes was mentioned in only 36% of our patients, confirming that valid studies of mortality in diabetes in the UK cannot be conducted using death certificates alone.

Errors Occurring in Linkage

The NHSCR holds 58 million records on residents in England and Wales in its computerized registration system. To match records on Diabeta successfully and accurately with the massive database in the NHSCR requires that the quality of data supplied to the NHSCR is very high. Insufficient discrimination power, noise, and missing data are major obstacles to the linkage of records.[10] The failure to find matches for 8.7 % of patient records in Diabeta may be due to inadequate or inaccurate demographic data recorded on Diabeta. Reasons for such inaccuracies include incomplete patient identification information in referral letters, as well as the relatively common problem of patients sharing the same names. Furthermore, demographic data (name, address) cannot be updated accurately over time, unless such changes are notified by the patient or GP. Despite notices in the waiting area, patients often forget to notify the receptionist about a change of address. In order to minimize this problem, the demographic screen appears automatically when a patient's record is opened. It is not infrequent that the patient sees this in the doctor consultation and comments that they haven't lived at that address for some time. In order to recover some records which cannot be linked with the NHSCR directly,

Matching procedure	Cost per record (£)	Number of traced records (%)	Total cost (£)
Automatically matching	0.30	2 877(42%)	863.1
Operator interference	1.70	2 672 (39%)	4 542.4
Manual matching	4.75	1 302 (19%)	6 184.5
Total		6 851	11 590.0

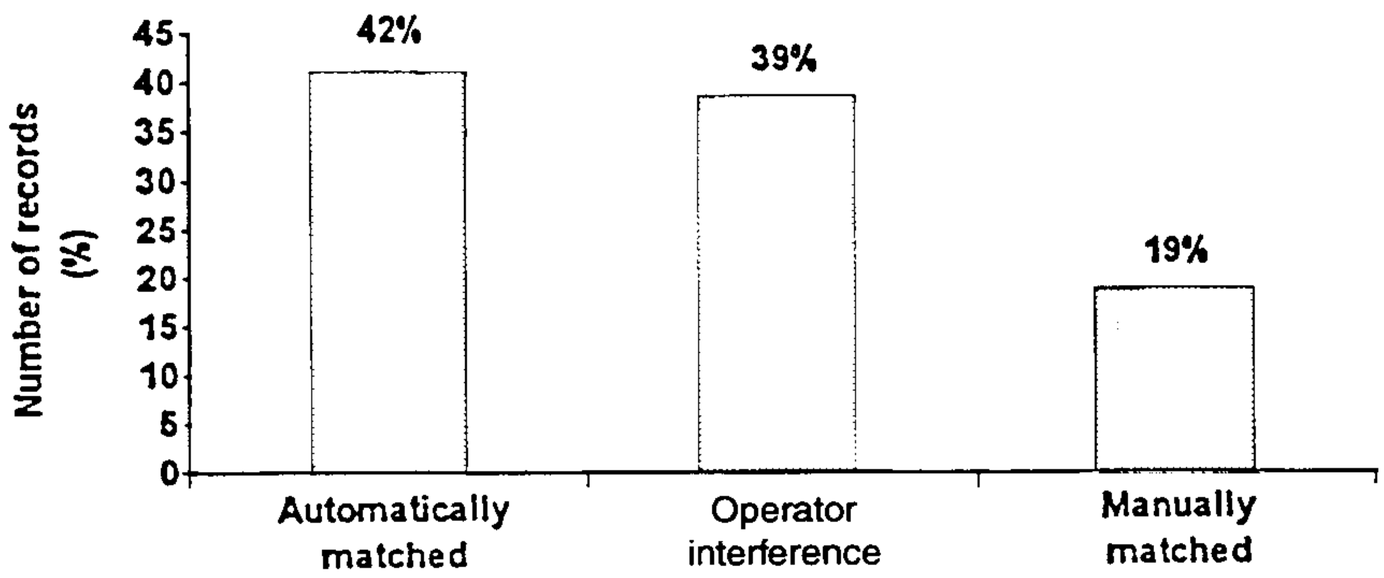


Figure 4. Procedures used in record matching

preliminary efforts at linkage of such records with a local FHSA may be helpful.

The FHSA registration database holds patient's address and GP details, and has better discriminating power than the NHSCR (e.g. it holds information on patient's current and previous address). Matching local patients' data in a local FHSA database is much quicker and more accurate than that performed in the NHSCR.

The registration database in the Lambeth, Southwark and Lewisham FHSA (LSL FHSA) covers most of the diabetic population in St Thomas' Hospital and may be linked to Diabeta. In one study, we selected a sample of 104 patients from Diabeta in order to evaluate the feasibility of implementing a linkage between Diabeta and the local FHSA. These records have been traced both in the NHSCR and in the LSL FHSA. Results showed that the local FHSA registration database enabled tracing of some patients who could not be traced by the NHSCR. In this sample, over 50% of records could not be fully matched in the NHSCR, but has been successfully matched in the LSL FHSA registration database. Patients' vital status and demographic data can be confidently updated in this way, so long as the patient remains resident in the LSL locality; if they have moved out of the area then the only source of linkage is through the NHSCR.

Achieving a Complete Update of Patient's Data on Diabeta in the Future

Extensive research is currently on-going in the Family Health Services Computer Unit (FHS CU) regarding the feasibility of installing DIALOG in each FHSA.[11] The DIALOG system is designed to interface with the FHSA registration database, enabling real-time downloading of demographic data.[12] It is intended to register all diabetic patients (with the patient's consent) within a district. Its main aim is to support annual check-ups. Based on this selection, a reminder could then be generated and sent to the lead clinician (consultant or GP). A standard clinical dataset[13] is collected from the lead clinician on a ASCII file diskette; this is then transferred onto DIALOG.

A national diabetes register is to be set up in the FHS CU. This register will hold registration on all diabetic patients in the NHS by linking the DIALOG system run in each FHSA.[14] Once this is available, any changes in patient demographic information reported to the FHSA will also be continually updated on the diabetes register. If patients move out of the FHSA area, this would be notified to the new FHSA and the patient's latest annual review report will be directed to the new FHSA.

A future link between Diabeta and DIALOG will eventually lead to a complete update of patient registration and clinical details in the Diabeta database in a cost-effective way. The cost for updating any diabetic clinical information systems is however quite substantial at about £3000 per annum.[12] DIALOG will help ensure that every diabetic patient has the opportunity to be reviewed on a regular basis. This linkage may make a substantial contribution towards reducing morbidity and mortality in our diabetic population.[14] Research in this area is however incomplete but pilot studies on the DIALOG system have already been performed in four FHSAs.[14]

Diabet. Med. 14: 877–883 (1997)

One further question arises: once DIALOG is introduced at a general level, will the link between Diabeta and the NHSCR still be useful? The answer is probably 'yes', as only the NHSCR can provide details on the cause of death in patients. The DIALOG system, on the other hand, will only provide the date and place of death and its forecast costs are substantially higher. Further developments on DIALOG in order to provide copies of death certificates are clearly indicated.

Information systems in hospital and community care are inadequate to monitor the target of St Vincent.[15] This study has introduced a method to improve the accuracy and completeness of patient data recorded in a hospital information system. It enables an information system to identify the true 'final outcome' of its diabetic population and to analyse the risk factors related to mortality in diabetic patients.[16] The evaluation of diabetes care based on hospital and community care can be misleading through inaccuracies in patients' vital status records in information systems. One study has been useful in tracing patients selected for long-term follow-up study related to foot complication.[17] Continual update of patients' vital status also prevents the unnecessary distress caused by sending appointments to patients who died. Lastly the NHS number, now recorded in Diabeta, can readily be used for transferring information within NHS information systems.

The feasibility of linking Diabeta with the NHSCR has resulted in a pilot trial. Implementation of this linkage has been very successful and we feel that it provides a stronger foundation for maintaining continuous and high quality patient care.

Acknowledgements

The study was supported by the British Council and the Directorate of Diabetes & Endocrinology within the Guy's and St Thomas' Hospitals' Trust. We are grateful to the Research Unit of the Office of Population Censuses and Surveys (OPCS) and also to all Family Health Services Authorities who were involved in this study and particularly to the very helpful people in the IT department of our local FHSA (Lambeth, Southwark and Lewisham).

References

1. Diabetes care and research in Europe. The St Vincent Declaration. *Diabetic Med* 1990; **7**: 360.
2. Baldwin JA, Achesoh ED, Graham WJ. *Text book of Medical Record Linkage*. Oxford: Oxford University Press, 1987; 6.
3. Newsletter of the Juvenile Diabetes Foundation (UK). *Diabetes Breakthrough* 1996; **9**: 3.
4. The newsletter for ONS' medical research services. *The Researcher*. Office for National Statistics 1996; **6**: 2.
5. Office of Population Censuses & Surveys. *Guidance Notes on Applications to Use the NHS Central Registers for Research*. London: OPCS, 1–7.
6. *Data Protection Act*. 1984. London: HMSO.
7. Thompson EJ. *Statement of Policies on Confidentiality and Security of Personal Data*. London: OPCS, 1993: 2.
8. Henderson J, Goldacre MJ, Griffith M, Simmons H. Recording of deaths in hospital information systems. Implications for audit and outcome studies. *Journal of Epidemiology & Community Health* 1992; **46**: 297–299.
9. Balkau B, Papoz L. Certification of cause of death in French diabetic patients. *Journal of Epidemiology and Community Health* 1992; **46**: 63–65.
10. Acheson ED. *Record Linkage in Medicine* Edinburgh and London: E. & S. Livingstone Ltd, 1968: 197–198.
11. Vaughan NJA, Shaw M, Boer F, Billett D, Martin C. Creation of a District Diabetes Register using the DIALOG system. *Diabetic Med* 1996; **13**: 175–181.
12. Vaughan NJA, Hopkinson N, Chishty VA. DIALOG: Co-ordination of the Annual Review Process through a District Diabetes Register linked to the FHSA database. *Diabetic Med* 1996; **13**: 182–188.
13. Vaughan NJA, Home PD for the Diabetes Audit Working Group of the Research Unit of the Royal College of Physicians and the British Diabetic Association. The UK Diabetes Dataset—a standard for information exchange. *Diabetic Med* 1995; **12**: 717–712.
14. *Diabetes Care Management. First Health*. East Sussex FHSA, Summer 1994, 6–7.
15. Sönksen PH, Harris A, Jarman B. Perceived strategic and practical problems in the use of information technology for quality improvement in diabetes care in the United Kingdom. *Diabetes & Metabolism (Paris)* 1993; **19**: 80–88.
16. Weng C, Coppini D, Mozzaka N, Sönksen PH. Deaths related to diabetic foot problems as an estimate of mortality associated with peripheral neuropathy (Abstract). *Diabetic Med* 1996; **13** (suppl 7) P79.
17. Coppini DV, Young PJ, Weng C, Sönksen PH. Risk factors associated with the development of end stage peripheral neuropathy: a 10–13 year follow-up study (Abstract). *Diabetic Med* 1996; **13** (Suppl 3) P85.